



ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACIÓN

Titulación:

INGENIERO DE TELECOMUNICACIÓN

Título del Proyecto:

SEPARACIÓN DE FUENTES SONORAS CUASI-
ARMÓNICAS

Iñigo Andiñón Iracheta
Miroslav Zivanovic

Pamplona, 20 de Julio de 2011

Índice

1. Introducción.....	5
2. Motivación.....	7
2.1 Problemática.....	8
2.2 Aplicaciones.....	9
2.3 Estado del arte.....	10
3. Objetivos	12
4. Análisis de señales procedentes de una fuente.....	15
4.1 Introducción	16
4.2 Modelado polinomial	17
4.3 Resultados del modelado polinomial	19
4.4 Estudio de los coeficientes del modelo.....	23
4.5 Filtro de Kalman	35
4.6 Resultados del filtro de Kalman.....	36
4.7 Estudio de los coeficientes del modelo.....	38
5. Separación de señales cuasi-armónicas procedentes de varias fuentes	42
5.1 Introducción	43
5.2 Modelado polinomial	48
5.3 Modelado polinomial con técnicas de interpolación	52
5.4 Filtro de Kalman	56
5.5 Filtrado en frecuencia.....	60
6. Estudio comparativo	65
6.1 Introducción	66
6.2 Impacto del ruido	68
6.3 Calidad de la estimación en función de la desviación de frecuencias fundamentales.....	75
7. Conclusiones y líneas futuras	78
7.1 Conclusiones	79
7.2 Líneas futuras.....	81
8. Bibliografía.....	83

1. Introducción

Las señales acústicas del mundo real normalmente están compuestas por la suma de señales procedentes de distintas fuentes sonoras, y llevar a cabo la descomposición de la mezcla en señales individuales puede llegar a ser una tarea complicada y un gran campo de investigación. Existen varios métodos para realizar la separación de fuentes sonoras, pero el rendimiento de los algoritmos existentes suele ser limitado, en comparación por ejemplo con el funcionamiento del oído humano, que es capaz de diferenciar fácilmente cada fuente individual presente en mezclas muy complejas.

El grado de éxito con el que se realiza la separación depende en gran medida de la cantidad de información previa que conozcamos al respecto. Es decir, si de antemano conocemos las características de las fuentes que componen la mezcla, será más fácil que identifiquemos correctamente cada señal. En este sentido, en este proyecto nos centraremos en estudiar métodos que no suponen que haya ninguna información previa conocida, aunque sí que haremos uso de las propiedades generales de las señales musicales, como por ejemplo el principio de continuidad frecuencial y temporal.

Otro factor importante de éxito es el formato con el que nos encontramos las señales que vamos a estudiar. Por ejemplo, una grabación realizada con varios micrófonos nos puede permitir realizar la localización espacial de cada fuente, y esto facilita enormemente la separación. Pero normalmente nos encontramos con señales estéreo (dos canales) o mono (un canal). En este proyecto trabajaremos con señales mono, por lo que no podremos aprovecharnos de ninguna ventaja de este tipo.

En estas circunstancias, podemos decir que trabajaremos en el escenario del peor caso posible.

Las limitaciones mencionadas anteriormente de los algoritmos existentes se ven acentuadas cuando las fuentes son cuasi-armónicas, es decir, cuando los espectros de las señales que conforman la mezcla tienen armónicos solapados. Cuando esto sucede, la mayoría de los métodos suelen dar resultados muy poco satisfactorios.

El presente proyecto va a estar enfocado en el estudio y comparación de distintos métodos existentes de separación de señales, estudiando en particular su comportamiento en presencia de armónicos fuertemente solapados en frecuencia, situación ante la cual se propondrán modificaciones de los algoritmos para intentar solucionar las deficiencias de los mismos.

2.Motivación

2.1 Problemática

Cuando se presentan varias fuentes sonoras simultáneamente, la señal $x(n)$ resultante se representa en el dominio temporal como la superposición de las señales $s_m(n)$, correspondientes a cada una de las fuentes que la componen:

$$x(n) = \sum_{m=1}^M s_m(n), \quad n = 1, \dots, N$$

donde $s_m(n)$ es la m -ésima fuente en el tiempo n , y M es el número de fuentes.

La separación de fuentes sonoras se define como la recuperación de una o varias fuentes $s_m(n)$ a partir de $x(n)$.

Hay diferentes formas de definir qué es una fuente sonora $s_m(n)$. En este proyecto consideraremos cada elemento físico que vibra, es decir, cada instrumento, como una fuente sonora.

Como hemos adelantado en la introducción, existen varios métodos para realizar la separación de las fuentes sonoras, pero el rendimiento de los algoritmos existentes suele ser bastante limitado, sobretodo en el caso de que se presenten armónicos fuertemente solapados en frecuencia. Se puede afirmar por tanto que el mayor problema a la hora de llevar a cabo la separación de fuentes sonoras es la presencia de fuentes cuasi-armónicas.

2.2 Aplicaciones

Centrándonos en las aplicaciones dentro del procesado de audio, queda claro que si tenemos una señal polifónica mezcla de varias fuentes, aplicar cualquier procesado solamente a alguna de las fuentes que la componen, individualmente y no al conjunto, resulta imposible sin la previa separación de señales. Sin embargo, si la separación de señales se realiza correctamente, sabiendo identificar cada fuente sonora, se abren multitud de aplicaciones de análisis y procesado de señales.

Dentro del campo del análisis de señales musicales, la aplicación más clara puede ser la transcripción automática de partituras. Esto se puede llevar a cabo mediante un pre-procesado que consistiría en la separación de fuentes sonoras, para después hacer un análisis individual de cada señal por separado. Identificando los cambios en el tiempo de amplitud y frecuencia fundamental (pitch) de cada instrumento musical, se podría llevar a cabo la transcripción.

Otra de las aplicaciones más claras, es la posibilidad de procesar independientemente cada una de las fuentes que componen una señal polifónica. Por ejemplo, se podría remezclar una canción (bajar o subir el volumen de ciertos instrumentos, aplicar efectos...) sin necesidad de tener todas las pistas que componen la mezcla. También se podría eliminar ciertos instrumentos, como las voces, para poder utilizar la canción en un karaoke. Este tipo de manipulaciones hasta ahora sólo eran posibles en la fase de producción de la canción, antes de llevar a cabo la mezcla.

Además de las mencionadas aplicaciones dentro de las señales de audio, los resultados de la separación de fuentes se pueden aplicar a cualquier tipo de señal, sin tener que ser necesariamente fuentes sonoras. Por ejemplo, se podrían utilizar los resultados desde el campo de la mecánica, para la detección de un mal funcionamiento en las vibraciones de un motor, hasta el campo de la medicina, para el diagnóstico de enfermedades, pudiendo analizar las señales de los aparatos médicos con gran precisión.

La base de todas estas aplicaciones, y de otras más, es la correcta identificación y separación de las fuentes que componen una señal, y por este motivo se sigue investigando en este campo.

2.3 Estado del arte

Los primeros métodos de separación de señales estaban enfocados al tratamiento de señales de voz, pero en los últimos años se ha incrementado el interés en las señales musicales haciendo que se desarrollen nuevos métodos en este campo. Los instrumentos musicales tienen mecanismos muy variados de producir fuentes sonoras, lo que hace que las características de las señales musicales tengan un amplio rango de variabilidad tanto en el dominio temporal como frecuencial.

Los métodos existentes de separación de fuentes sonoras suelen dividirse en las tres siguientes categorías, aunque puede haber métodos que combinen aspectos de varias de ellas:

1. Métodos basados en el modelado sinusoidal.

Este tipo de métodos modelan las señales como un conjunto de sinusoides estables, llamadas parciales, más un ruido o componente residual. Es decir, se describe la señal para una trama $x(n)$ de la siguiente manera:

$$x(n) = \sum_{h=1}^H a_h \cos(2\pi f_h n / f_s + \theta_h) + r(n), \quad n = 0, \dots, N-1,$$

donde n es el índice temporal, N el número de muestras de la trama, f_s es la frecuencia de muestreo, a_h , f_h y θ_h son respectivamente la amplitud, frecuencia y fase inicial de la parcial h -ésima, y $r(n)$ es el ruido.

De esta manera, se modelan las señales de audio como la suma entre una serie de sinusoides y un ruido residual. La parte sinusoidal o determinista corresponde a los modos de vibración del sistema, mientras que la parte residual o estocástica recoge la energía producida por los mecanismos de excitación y otras componentes energéticas de naturaleza no sinusoidal.

Este tipo de métodos detectan y estiman los parámetros armónicos y suele ser muy utilizado también en procesamiento de la señal. Un paso crítico consiste en detectar las frecuencias fundamentales f_0 , ya que después la señal suele sufrir modulaciones en frecuencia (FM) y amplitud (AM), además de la suma de ruido aditivo. Una vez que se ha detectado f_0 se suelen estimar el resto de parámetros mediante mínimos cuadrados (LS), tras solucionar ecuaciones lineales tratando de minimizar el error entre la señal original y la estimada.

Dos de los métodos estudiados en este proyecto pertenecen a esta categoría, concretamente el basado en el modelado polinomial y el filtro de Kalman.

2. Métodos basados en el aprendizaje no supervisado.

Este tipo de métodos utilizan un modelo no paramétrico, y no tiene en cuenta ningún tipo de información relativa a las propiedades generales de las señales musicales, al

contrario que la anterior categoría. Este tipo de métodos describe las señales x_t como una suma ponderada de funciones base b_j , de la siguiente manera:

$$\hat{x}_t = \sum_{j=1}^J g_{j,t} b_j \quad t = 1, \dots, T,$$

Donde J es el número de funciones base b_j , y $g_{j,t}$ es la ganancia de la función base j -ésima en la trama t -ésima. También se puede expresar de la misma forma de manera matricial:

$$\hat{\mathbf{X}} = \mathbf{B}\mathbf{G},$$

Donde $\mathbf{X} = [x_1, x_2, \dots, x_T]$ es la matriz de observación, que representa lo que ha sido observado en T tramas; $\mathbf{B} = [b_1, b_2, \dots, b_J]$ es la matriz de funciones base, y \mathbf{G} es la matriz de ganancias donde $[G]_{j,t} = g_{j,t}$.

Las mencionadas funciones base se obtienen de la factorización del espectrograma en magnitud, en componentes fijas de frecuencia, y las ganancias son variables en el tiempo.

Existen diferentes técnicas desarrolladas para este tipo de métodos, algunas de ellas se basan en la independencia estadística de las fuentes, en la no negatividad de las matrices \mathbf{B} y \mathbf{G} , o en minimizar el número de elementos de las matrices \mathbf{B} y \mathbf{G} .

3. Métodos basados en características psicoacústicas.

Existen también modelos computacionales, que primero descomponen las señales en sus componentes elementales tiempo-frecuencia, y después las agrupan en sus respectivas fuentes sonoras. Las claves para asociar correctamente cada componente con su correspondiente fuente se basan en los mismos mecanismos que utiliza el cerebro humano para llevar a cabo la separación de fuentes, concretamente según Albert S. Bregman:

- Proximidad espectral (cercanía en tiempo o frecuencia)
- Concordancia armónica
- Cambios sincrónicos de los componentes (ataque común, caída común...)
- Proximidad espacial

3 Objetivos

Después de lo expuesto en el apartado de Motivación, queda clara la necesidad de realizar la separación de señales de forma correcta, incluso para armónicos fuertemente solapados en frecuencia, puesto que los métodos existentes no proporcionan una solución satisfactoria en estas circunstancias.

De esta manera, el principal objetivo de este proyecto es el desarrollo y comparación de distintos métodos de separación de señales cuasi-armónicas, mediante el estudio y comparación de tres métodos alternativos. El primero se basa en la estimación de los parámetros del modelo polinomial mediante técnicas de interpolación, el segundo se basa en el uso del filtro de Kalman, y el tercero se basa en el filtrado espectral.

Los pasos necesarios que desarrollaremos a continuación para conseguir el objetivo principal son los siguientes:

PRIMERA PARTE:

Señales procedentes de una sola fuente

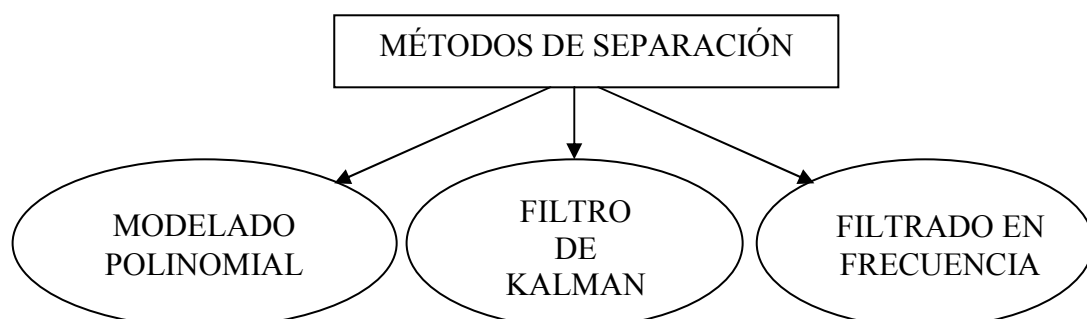
Estudio y estimación de señales cuasi-estacionarias procedentes de una sola fuente y evaluación de los algoritmos correspondientes al modelado polinomial y al filtro de Kalman. Se estudiará para cada tipo de señal:

- Señales originales y sus correspondientes estimaciones, haciendo un análisis por tramas, en dominio temporal y frecuencial.
- Obtención de los coeficientes del modelo.
- Evolución temporal de dichos coeficientes.
- Estimación de coeficientes mediante técnicas de interpolación: Obtención de los coeficientes de las tramas que presenten armónicos solapados a través de la información de las tramas vecinas.

SEGUNDA PARTE:

Separación de señales cuasi-armónicas procedentes de varias fuentes

Extrapolación de los resultados de la primera parte del PFC al estudio de señales más complejas, compuestas por señales procedentes de varias fuentes y con armónicos fuertemente solapados en frecuencia. Llevar a cabo la separación de señales mediante los siguientes métodos:



- **Modelado polinomial:**

El modelado polinomial describe las señales como la suma de estos dos componentes:

- Parte determinista.
- Ruido aditivo.

Para realizar una buena aproximación será especialmente importante la correcta detección inicial de los picos espectrales (parte determinista), ya que después variarán en el tiempo sufriendo modulación en amplitud y frecuencia. Una vez que se detecta la frecuencia fundamental de cada señal se trata cada armónico de manera individual y se estiman el resto de parámetros mediante mínimos cuadrados.

- **Filtro de Kalman:**

La descripción de las señales en este caso también será la suma de una parte determinista y otra de ruido, pero en este caso tendremos dos tipos de ruido:

- Ruido de medida: Debido al error de observación.
- Ruido proceso: Debido al error del modelo.

El buen funcionamiento de este algoritmo depende en gran medida de la correcta elección de los valores iniciales del ruido de medida y el ruido de proceso. Se realizará un estudio en relación a la elección estos valores y se estudiarán posibles modificaciones del algoritmo para introducir mejoras.

- **Filtrado en frecuencia:**

El método que sigue este algoritmo consiste básicamente en filtrar el conjunto de armónicos que se solapan en cada trama, dividiendo el pico espectral donde se produce el solapamiento para asignar cada parte a su fuente correspondiente.

Para cada uno de los métodos se obtendrán:

- Espectrogramas de la señal original compuesta.
- Señales originales y sus correspondientes estimaciones tras la separación, haciendo un análisis por tramas, en dominio temporal y frecuencial.
- Obtención de los coeficientes del modelo.
- Evolución temporal de dichos coeficientes.
- Impacto de modulaciones de amplitud y frecuencia (efectos de trémolo y vibrato).
- Impacto del ruido.

TERCERA PARTE:

Estudio Comparativo

Se realizará una comparación de los resultados obtenidos con los distintos métodos mediante el estudio de indicadores como la relación señal a residuo (SRR). A partir de los resultados se extraerán las conclusiones oportunas.

4 Análisis de señales procedentes de una fuente

4.1 Introducción

En este apartado se presenta el estudio realizado acerca de la estimación de señales cuasi-estacionarias procedentes de una sola fuente mediante los algoritmos correspondientes al modelado polinomial y al filtro de Kalman.

Durante la realización del proyecto se han utilizado señales reales de distintos instrumentos musicales obtenidas de una grabación en directo. Las características de estas señales son las siguientes:

- 44100 Hz de frecuencia de muestreo
- Codificación PCM de 8 bits por muestra
- 1 canal (señal mono)
- Ruido convolutivo del propio instrumento
- Ruido aditivo (de fondo) debido al ruido ambiente de la grabación

A continuación se mostrará el comportamiento de los algoritmos mencionados, concretamente la estimación de señales de flauta y de trompeta, haciendo un análisis por tramas en dominio temporal y frecuencial. Se ha hecho un estudio con 10 armónicos.

Además de estas características técnicas, supondremos que las señales utilizadas llevan asociadas modulaciones de amplitud y frecuencia (AM+FM). Muchas situaciones se modelan de esta manera, como por ejemplo la presencia de ‘vibrato’ en la señal, que puede definirse como una AM inducida por la FM. Gracias a esto podremos representar las variaciones en amplitud, frecuencia y fase que sufren las señales, ya que a pesar de tratarse de señales cuasi-estacionarias, es inevitable la presencia de pequeñas variaciones temporales. De esta manera, podemos representar la señal de test que utilizaremos así:

$$s_t(n) = \sum_{i=1}^{10} A_i(n) \cos[2\pi i F_0 n + \varphi_i(n)] + r(n),$$

$$\varphi_i(n) = i A_{FM} \sin(2\pi F_{FM} n + \alpha) + \gamma_i,$$

$$A_i(n) = \frac{A_0}{i} [1 + A_{AM} \cos(2\pi F_{AM} n + \delta)].$$

Las tasas de modulación F_{AM} , F_{FM} , A_{AM} y A_{FM} son seleccionados de tal forma que se asegure que el pico sea detectado correctamente con el lóbulo principal de la STFT, o lo que es equivalente:

$$F_{AM} = 2 F_{FM}; A_{AM} = 0,5; A_{FM} = 2$$

4.2 Modelado polinomial

El primer paso del análisis consiste en estudiar el comportamiento del algoritmo correspondiente al modelado polinomial. Como hemos adelantado en el apartado de Motivación, el modelado sinusoidal descompone las señales como un conjunto de sinusoides estables, llamadas parciales, más un ruido o componente residual. Es decir, se describe la señal para una trama de tamaño N de la siguiente manera:

$$s(n) = \sum_{i=1}^I A_i(n) \cos[2\pi i F_0(n)n + \varphi_i] + r(n), \quad n = 0, 1, \dots, N-1.$$

donde la parte determinista se corresponde con la superposición de I componentes armónicos, llamados parciales, con amplitudes y frecuencias variantes en el tiempo; mientras que la componente residual es típicamente una secuencia de ruido gaussiano o coloreado. La frecuencia F_0 se expresa en ciclos/muestra y la fase φ_i en radianes.

Aplicando identidades trigonométricas, podemos expresar la ecuación anterior de la siguiente forma:

$$s(n) = \sum_{i=1}^I a_i(n) \sin(2\pi i F_0(n)n) + b_i(n) \cos(2\pi i F_0(n)n) + r(n),$$

$$a_i(n) = -A_i(n) \sin \varphi_i, \quad b_i(n) = A_i(n) \cos \varphi_i.$$

Los parámetros $a_i(n)$, $b_i(n)$, y $F_0(n)$ son funciones tiempo-variantes, que pueden ser modeladas como polinomios o una suma de funciones base. Al tratarse de señales musicales, asumiremos que la amplitud y la frecuencia fundamental tendrán variaciones continuas a lo largo de la ventana de análisis, y usaremos las siguientes aproximaciones polinómicas para el modelo:

$$\hat{a}_i(n) = \sum_{m=0}^M a_m^{(i)} n^m, \quad \hat{b}_i(n) = \sum_{m=0}^M b_m^{(i)} n^m,$$

$$\hat{F}_0(n) = \sum_{m=0}^M F_m n^m.$$

Por simplicidad, supondremos que los polinomios tienen el mismo orden M . Combinando las ecuaciones anteriores, y operando matemáticamente, podemos llegar a reescribir la ecuación del modelo sinusoidal como la siguiente expresión, que utilizaremos en este modelo polinomial:

$$s(n) = \sum_{i=1}^I p_s^{(i)}(n) \sin(2\pi i F_0 n) + p_c^{(i)}(n) \cos(2\pi i F_0 n) + r(n),$$

$$p_s^{(i)}(n) = \sum_{k=0}^{2M+1} \left[a_k^{(i)} - 2\pi i \sum_{m=0}^{k-2} b_m^{(i)} \frac{F_{k-m-1}}{k-m} \right] n^k,$$

$$p_c^{(i)}(n) = \sum_{k=0}^{2M+1} \left[b_k^{(i)} + 2\pi i \sum_{m=0}^{k-2} a_m^{(i)} \frac{F_{k-m-1}}{k-m} \right] n^k,$$

El primer paso para realizar la estimación de señales utilizando este modelo polinomial es detectar la frecuencia fundamental F_0 . Será crítico realizar una buena primera estimación de F_0 , ya que después la señal suele sufrir modulaciones en frecuencia (FM) y amplitud (AM), además de la suma de ruido aditivo. El método elegido para realizar esta tarea es el análisis espectral automatizado ('ASA': automated spectral analysis), debido a que la estimación de F_0 se realiza independientemente del resto de parámetros del modelo. El funcionamiento del método 'ASA' consta de dos pasos:

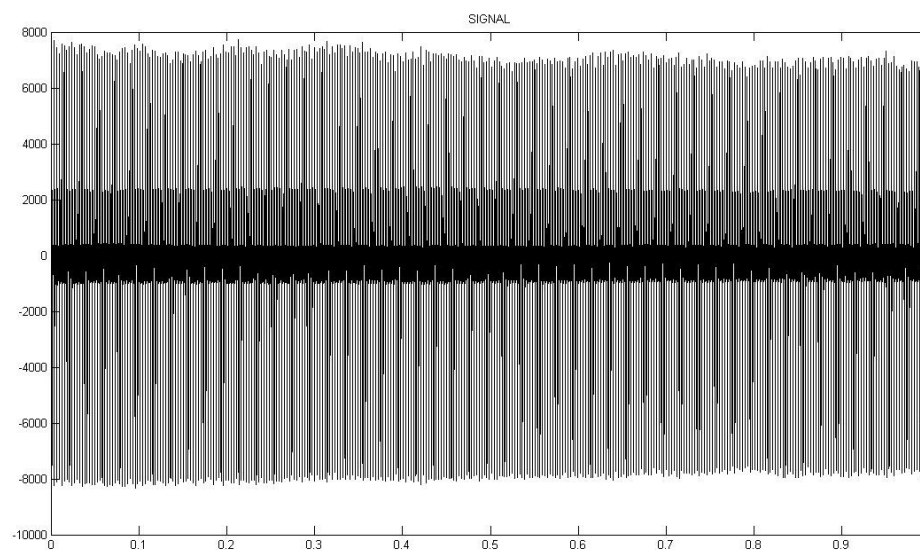
1. **Estimación de F0 inicial:** basada en el método de correlación, que detecta la distancia entre picos sucesivos en la secuencia de autocorrelación de la señal de audio. Si el espectro es plano y de banda ancha la estimación es correcta, pero si es de banda estrecha, esta estimación inicial falla y necesita ser depurada.
2. **Estimación de F0 mejorada:** La ventana de análisis recoge varios periodos de la señal $s(n)$, aunque la señal no es estrictamente periódica porque sus parámetros armónicos son tiempo-variantes. El espectro detectado sufre el efecto 'leakage', debido al análisis DFT, que es mitigado mediante interpolación parabólica. La estimación final de F_0 converge rápidamente tras pocas iteraciones del algoritmo.

Una vez que tenemos el valor de la frecuencia fundamental F_0 , el resto de parámetros del modelo (los coeficientes $p_s^{(i)}$ y $p_c^{(i)}$ de la ecuación anterior) se obtienen mediante el método de mínimos cuadrados (LS).

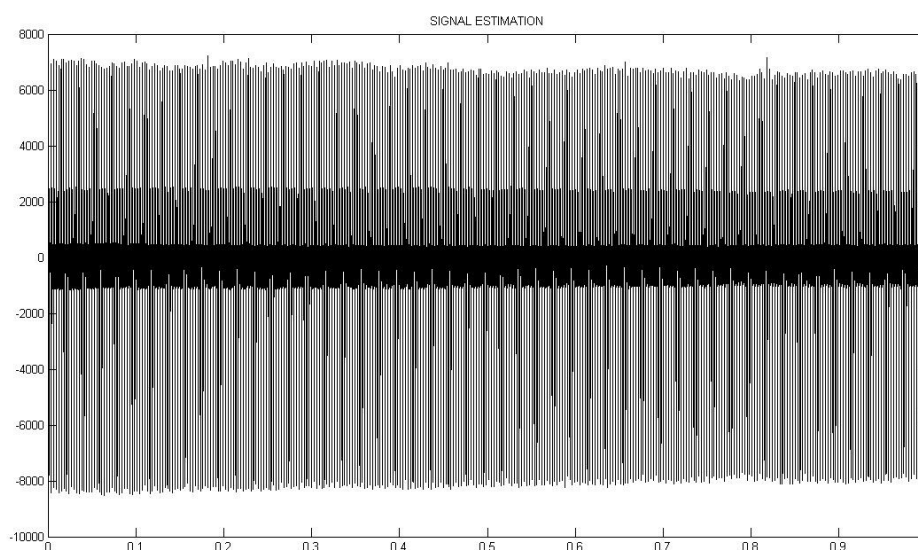
4.3 Resultados del modelado polinomial

A continuación se muestra el comportamiento del modelo polinomial, concretamente la estimación de una señal de flauta de 1 segundo de duración, haciendo un análisis por tramas de 500 muestras, con un solapamiento entre tramas del 50%. Se han utilizado polinomios de orden 4, y se ha hecho un estudio con 10 armónicos. Como es lógico, a mayor número de armónicos y mayor orden en los polinomios, se obtienen mejores resultados, pero también acarrea mayor coste computacional.

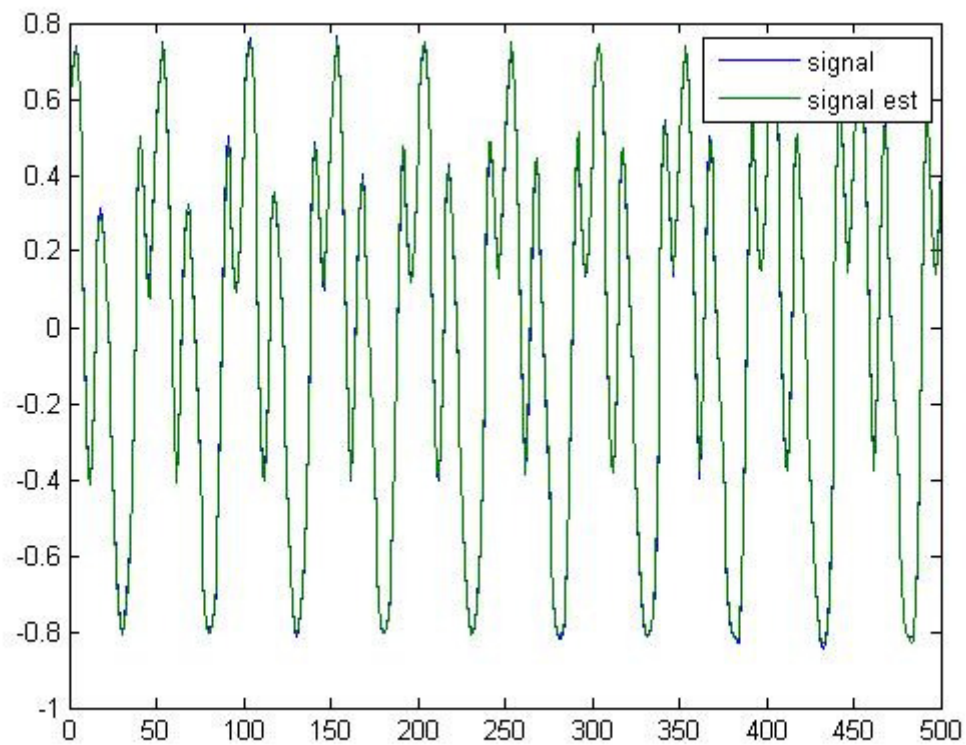
Señal original en dominio temporal:



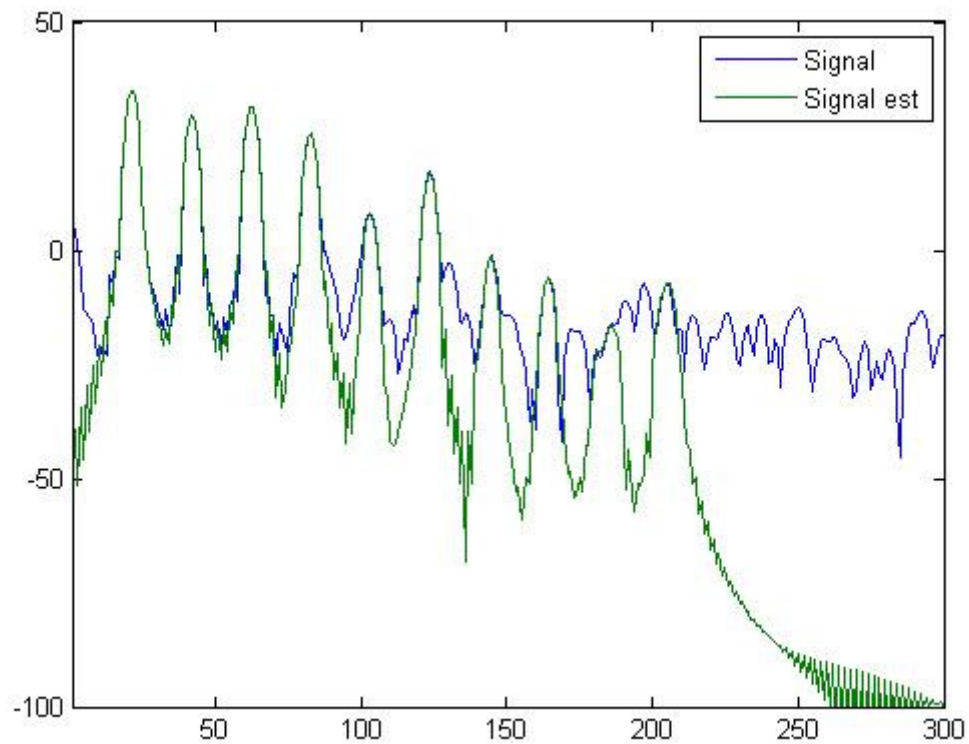
Estimación de la señal en dominio temporal:



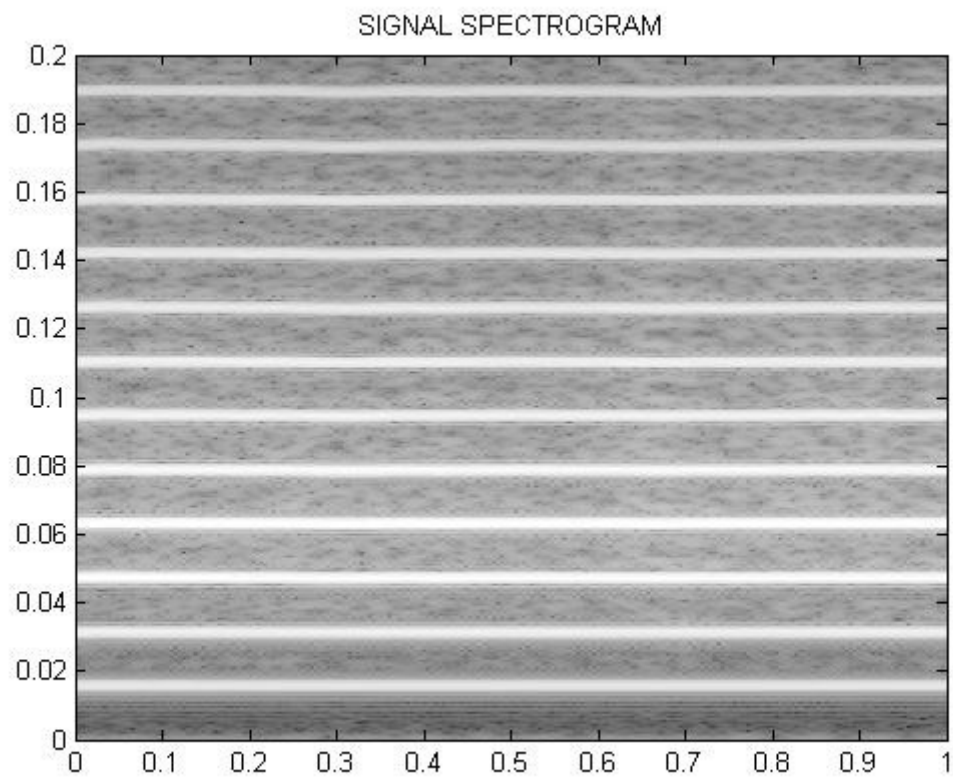
Trama de 500 muestras, señal original normalizada y estimación en dominio temporal:



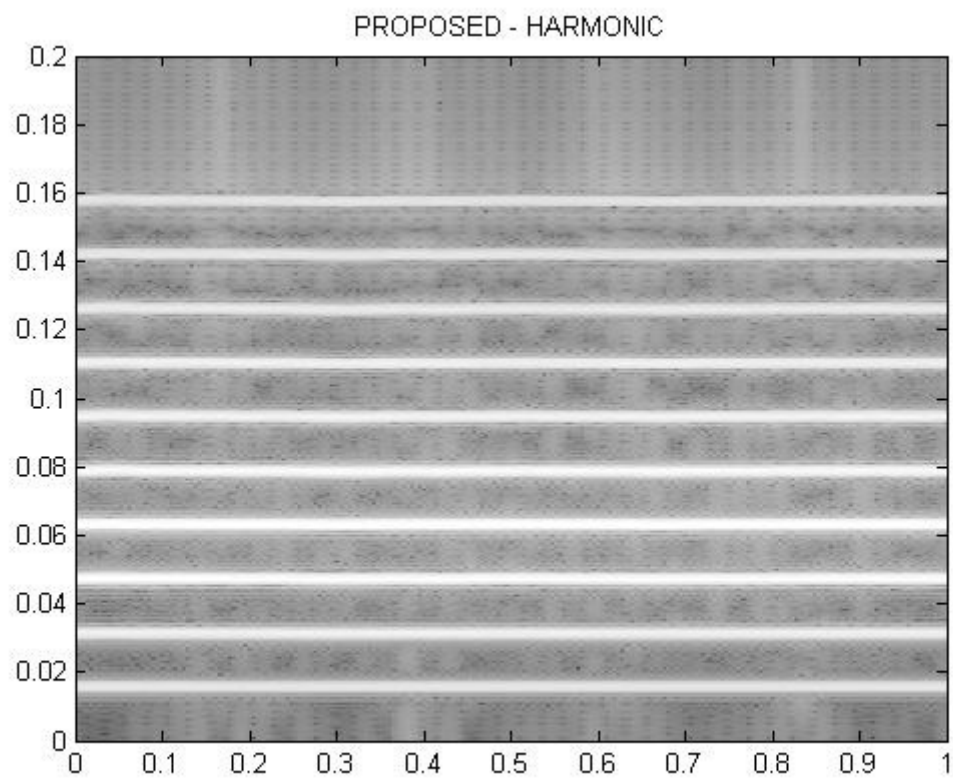
Señal original y estimación en dominio frecuencial:



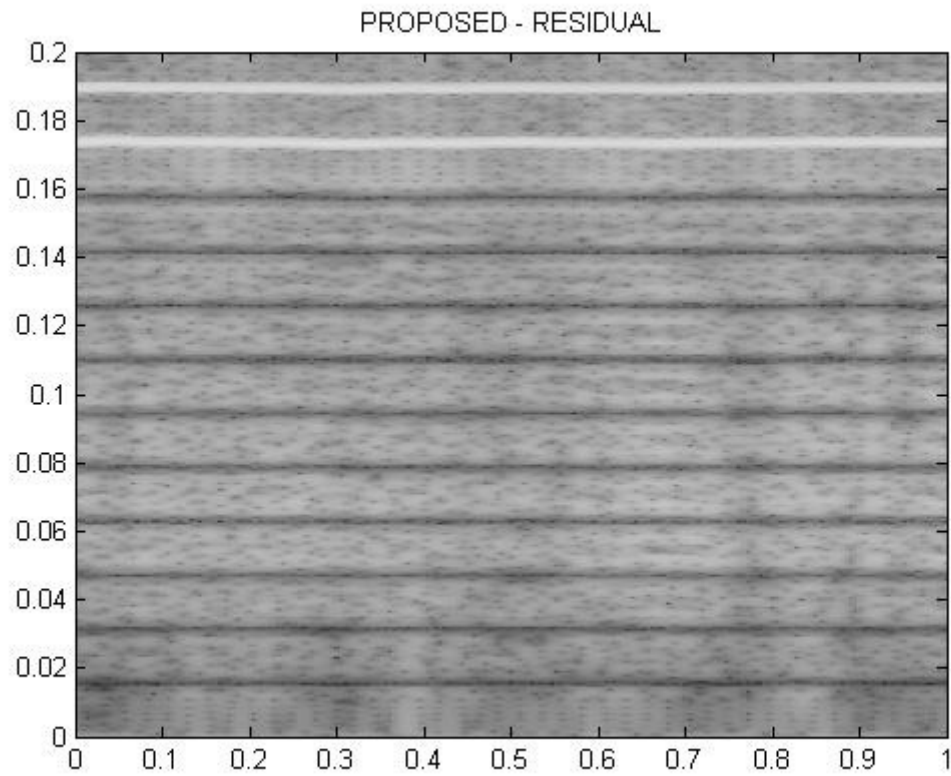
Espectrograma de la señal original:



Espectrograma de la señal estimada:



Espectrograma de la diferencia entre la señal original y la señal estimada:



4.4 Estudio de los coeficientes del modelo

Para comprender mejor el comportamiento del modelo polinomial se ha realizado un estudio de los coeficientes del modelo analizando distintas situaciones, que se desarrollarán a continuación:

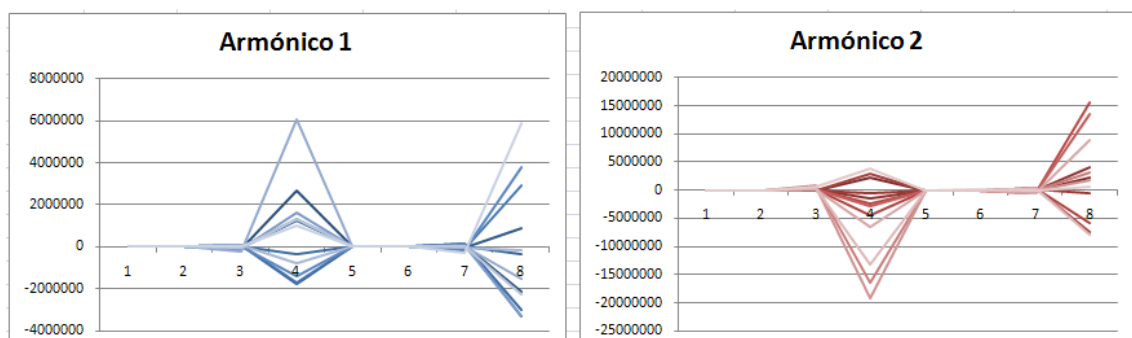
- Análisis de los coeficientes de un solo armónico en muchas tramas
- Análisis del mismo coeficiente para todos los armónicos en muchas tramas
- Análisis de los coeficientes de todos los armónicos menos uno, y deducción del valor de los coeficientes desconocidos a partir de los armónicos vecinos mediante interpolación lineal
- Estudio del criterio de continuidad temporal de los coeficientes para una fuente sonora cuasi-armónica
- Modificación de la base de tiempos del modelo

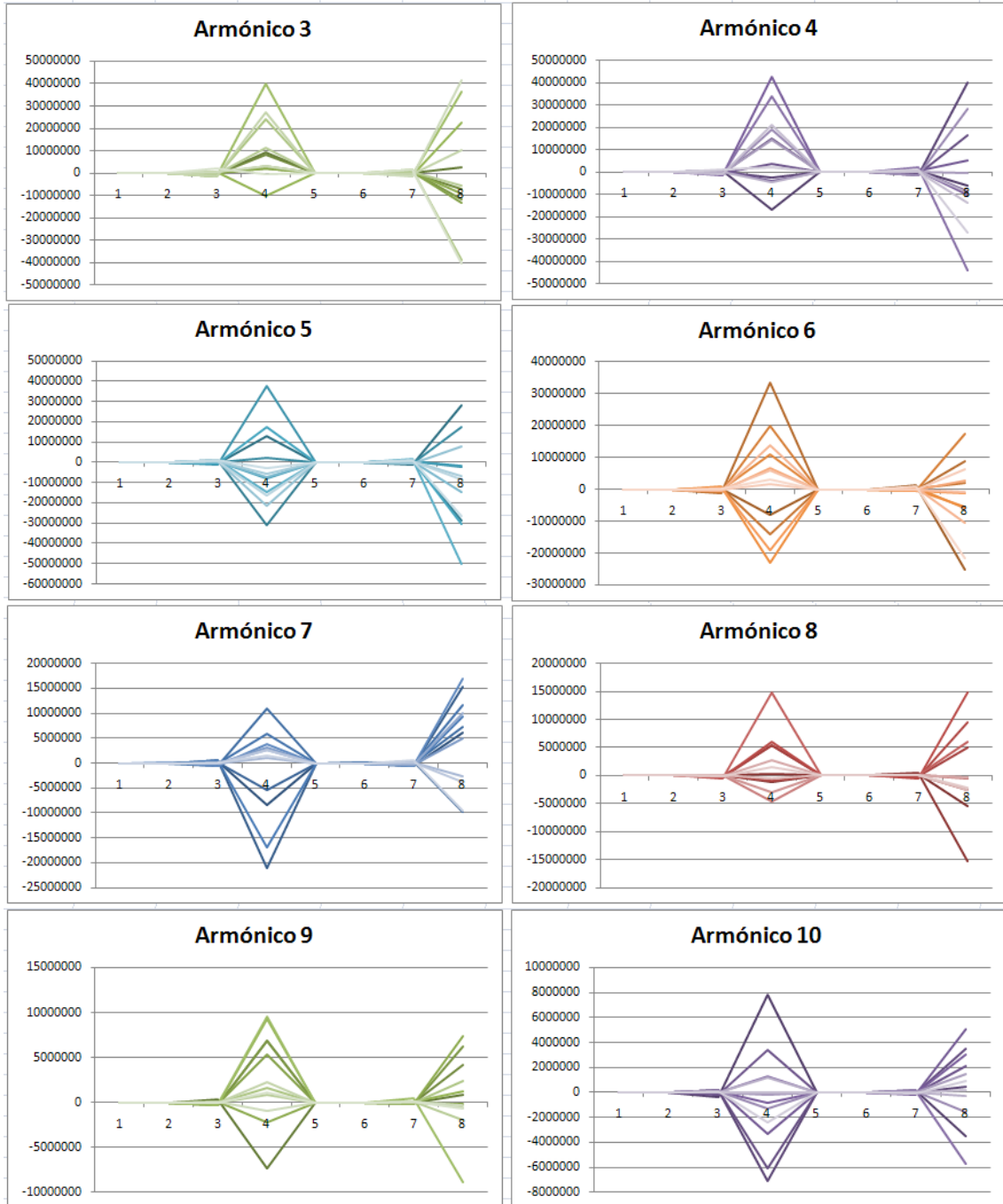
4.4.1 Análisis de los coeficientes de un solo armónico en muchas tramas:

En las siguientes gráficas se muestra en el eje x los valores de los 8 coeficientes del modelo correspondientes a un análisis con polinomios de orden $M=4$, lo que significa que los 4 primeros son los que multiplican al seno (a_i) y los 4 últimos los que multiplican al coseno (b_i), retomando la expresión del modelo:

$$s(n) = \sum_{i=1}^I a_i(n) \sin(2\pi i F_0(n)n) + b_i(n) \cos(2\pi i F_0(n)n) + r(n),$$

$$\hat{a}_i(n) = \sum_{m=0}^M a_m^{(i)} n^m, \quad \hat{b}_i(n) = \sum_{m=0}^M b_m^{(i)} n^m,$$





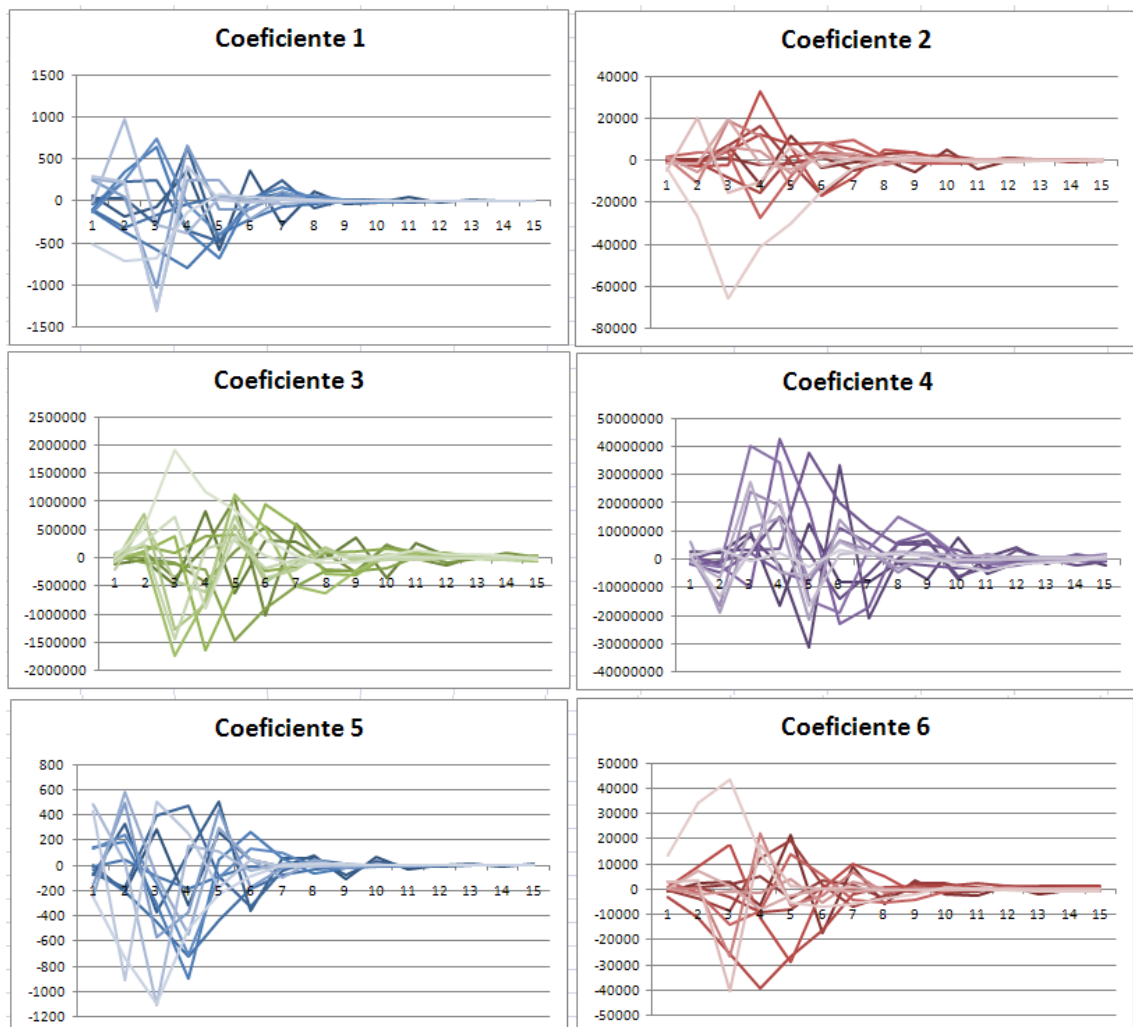
Tras observar las gráficas anteriores queda claro que los coeficientes de mayor orden (4 y 8) adquieren valores mayores. Aunque esto no quiere decir que estos coeficientes tengan mayor peso en la contribución a la estimación, sino todo lo contrario, puesto que los coeficientes de mayor orden multiplican a potencias superiores de n , y n es la base de tiempos que toma valores cercanos a cero. Por lo tanto este comportamiento es lógico y esperable. Para un armónico concreto tenemos:

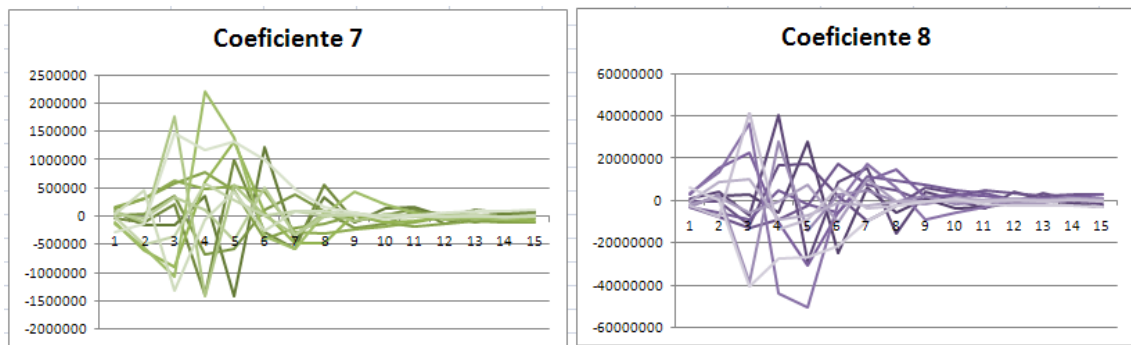
$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\text{sen } 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\text{cos } 2\pi f_0n$$

donde n es la base de tiempos que adquiere los valores: $t = (0:N-1)$, donde N es el tamaño de la ventana de análisis.

4.4.2 Análisis del mismo coeficiente para todos los armónicos en muchas tramas:

En las siguientes gráficas se muestran los valores de un coeficiente para todos los armónicos, en muchas tramas. Esta vez se ha realizado un análisis con 15 armónicos, que se muestran en el eje x:



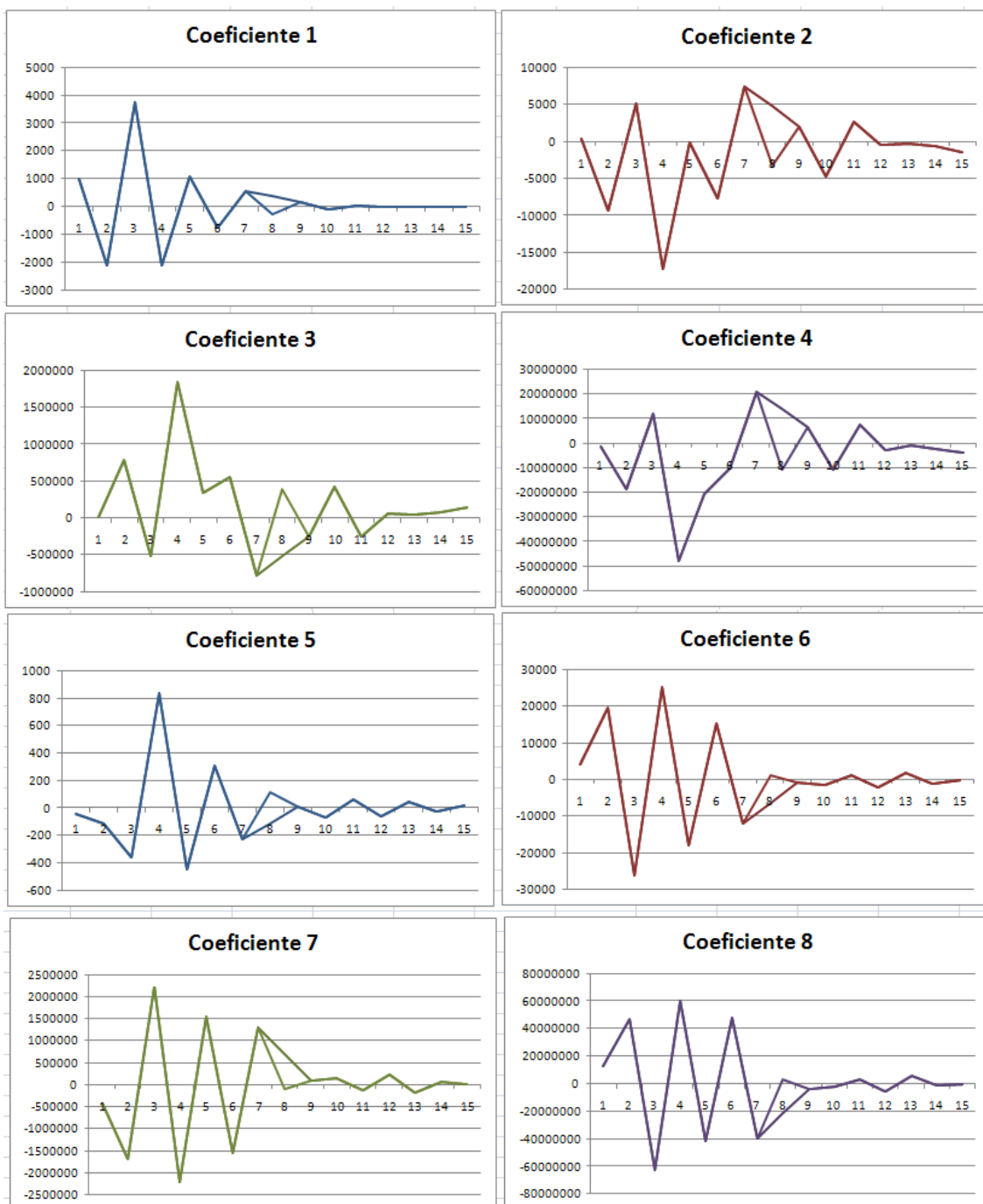


Como era de esperar, los armónicos superiores adquieren valores más pequeños, puesto que en el espectro de la señal, los armónicos correspondientes a frecuencias más altas tienen menos energía.

4.4.3 Análisis de los coeficientes de todos los armónicos menos uno, y deducción del valor de los coeficientes desconocidos a partir de los armónicos vecinos mediante interpolación lineal.

Una vez que hemos visto los valores que suelen tomar los coeficientes del modelo, intentaremos abordar la problemática que surge cuando los coeficientes de cierto armónico son desconocidos. Esta situación nos sirve para modelar el caso de que intentemos separar dos señales que tienen algún armónico solapado en frecuencia, pero no el resto.

Las gráficas que se muestran a continuación son similares a las anteriores, con la diferencia de que sólo se muestran los coeficientes de una trama. Se han obtenido los coeficientes para el armónico 8, que se supone desconocido, y en las figuras se muestra el valor que se obtendría a través del algoritmo, y el valor obtenido por interpolación lineal a partir de los valores de los armónicos vecinos:



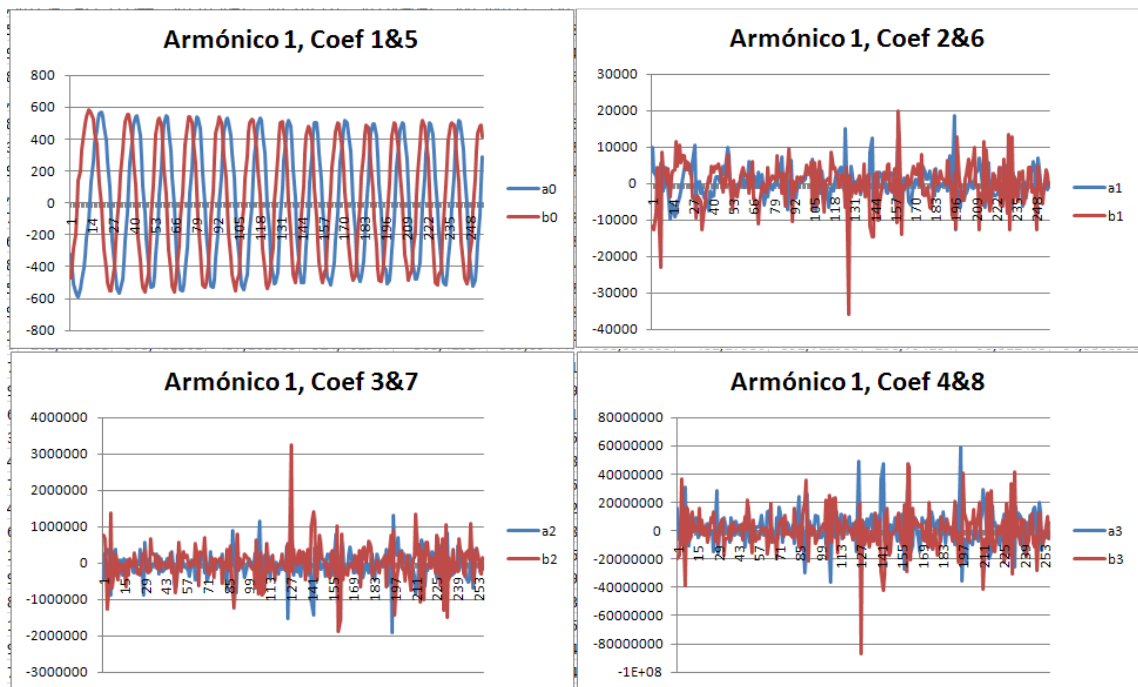
Al observar las gráficas a primera vista, parece ser que deducir el valor de los coeficientes a través de la interpolación lineal entre armónicos vecinos, no produce un buen resultado, por lo que de momento desechamos la idea de aproximar el valor de un coeficiente a través de los armónicos vecinos.

4.4.4 Estudio del criterio de continuidad temporal de los coeficientes para una fuente sonora cuasi-estacionaria

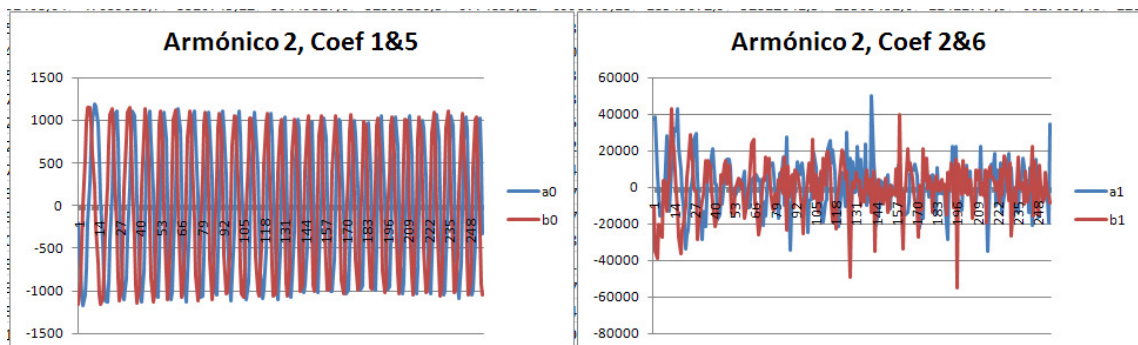
Para comprender el comportamiento de los coeficientes del modelo, resulta muy importante estudiar su evolución temporal. En las gráficas siguientes se observa cómo varían los valores de los coeficientes al avanzar las tramas en el tiempo. Se ha utilizado un solapamiento del 50% entre tramas, y se muestran los coeficientes emparejados de la siguiente manera: (a_0, b_0) ; (a_1, b_1) ; (a_2, b_2) ; (a_3, b_3) ;

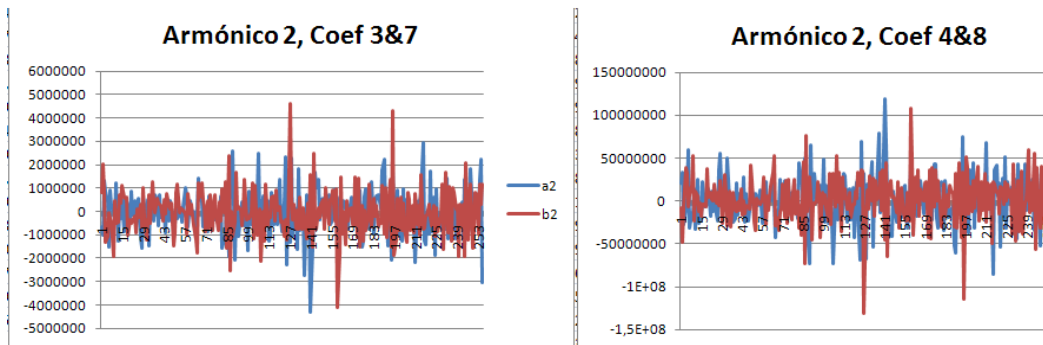
$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\text{sen } 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\text{cos } 2\pi f_0n$$

Armónico 1:

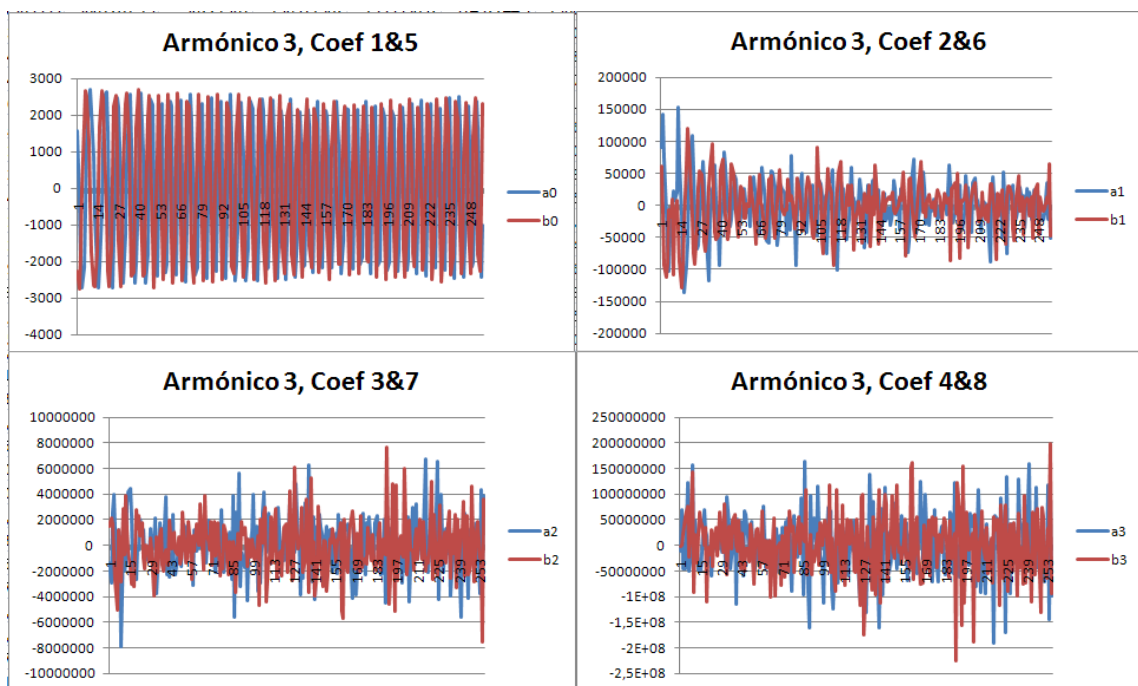


Armónico 2:





Armónico 3:



Si continuamos observando los coeficientes correspondientes al resto de armónicos, podemos ver que se comportan siguiendo el mismo patrón:

- los coeficientes (a_0 y b_0) siguen un patrón periódico conforme avanzan las tramas, similar al de las funciones seno y coseno. Desfasados entre sí y con frecuencia mayor cuanto más alto es el valor del armónico.
- el resto de coeficientes toman valores más impredecibles, aunque dentro de un rango.

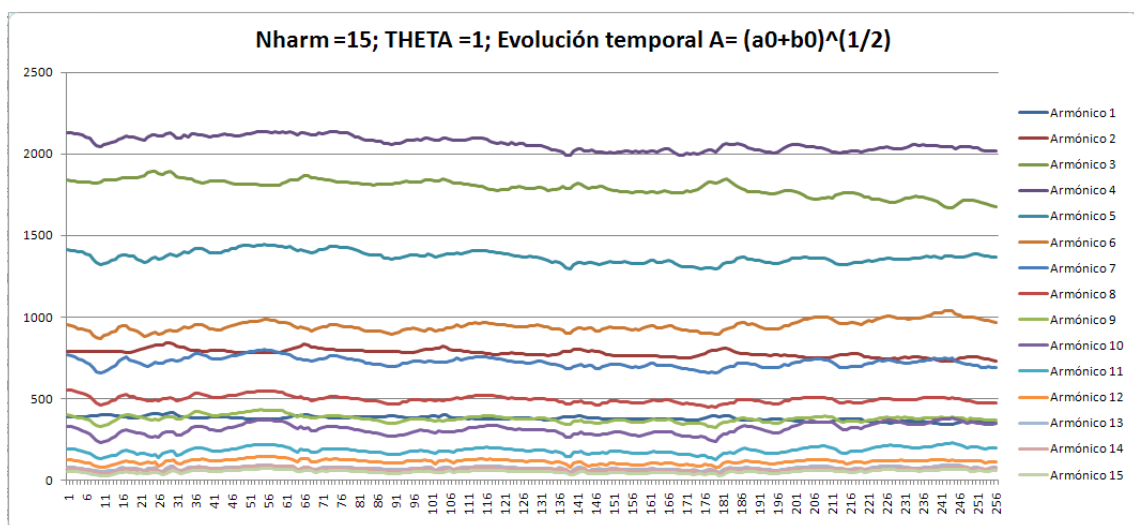
Al tratarse de una señal cuasi-estacionaria, a priori cabría esperar que los coeficientes se mantuvieran casi constantes en el tiempo, o por lo menos que tuvieran poca variabilidad, pero hemos visto que esto no sucede así.

Esto es debido a que la amplitud de la señal para cada armónico no está representada por los coeficientes directamente, sino por la combinación de ellos. Por ejemplo, si aplicamos el modelo con polinomios de orden $M=1$, para cada armónico tendríamos:

$$s(n) = (a_0)\sin 2\pi f_0 n + (b_0)\cos 2\pi f_0 n$$

y la amplitud de cada armónico quedaría representada por $A = \sqrt{a_0^2 + b_0^2}$.

Si observamos la evolución temporal de este coeficiente $A = \sqrt{a_0^2 + b_0^2}$ observamos que sí que se mantiene casi constante al avanzar las tramas en el tiempo:

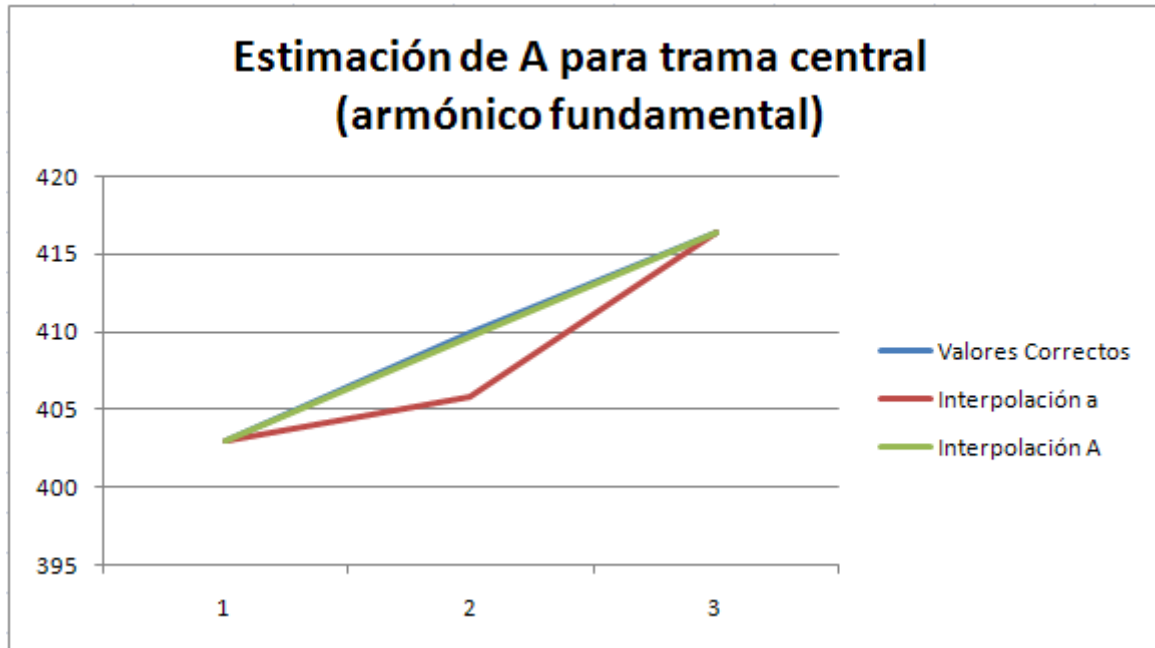


Sin perder de vista el objetivo principal de este proyecto, que es resolver el solapamiento de armónicos, y tras observar la evolución temporal de los coeficientes del modelo, nos planteamos la posibilidad de aproximar los valores de ciertas tramas a partir de la información de las tramas vecinas. Aprovechándonos del criterio de continuidad temporal y la suavidad espectral con la que varían las señales musicales, intentaremos deducir el valor de los coeficientes desconocidos a partir de las tramas vecinas mediante interpolación lineal.

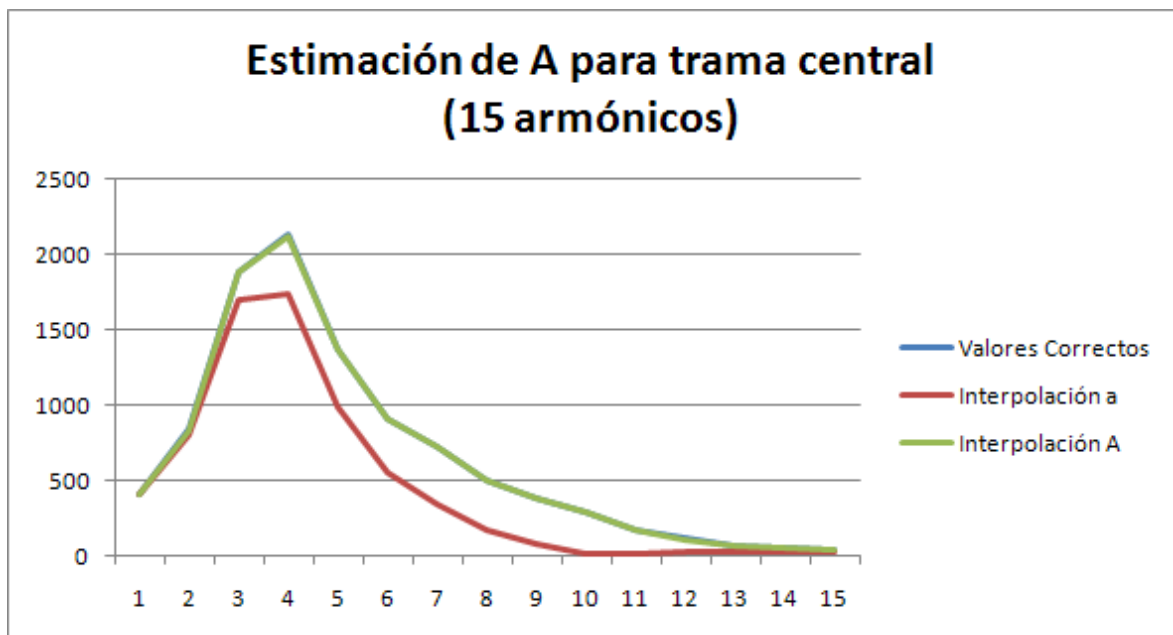
Las siguientes gráficas verifican el criterio de continuidad temporal. Utilizando un solapamiento entre tramas del 50%, escogiendo 3 tramas adyacentes se deduce el valor de $A = \sqrt{a_0^2 + b_0^2}$ de la trama intermedia por interpolación lineal, y se observa la poca variación que hay entre el valor real y el interpolado.

Esta gráfica muestra el valor de A en el eje y, y el número de trama en el eje x:

Esta gráfica muestra el valor de A en el eje y, y el número de trama en el eje x:



Y aquí se muestra el valor de A en el eje y, y el número de armónico en el eje x:

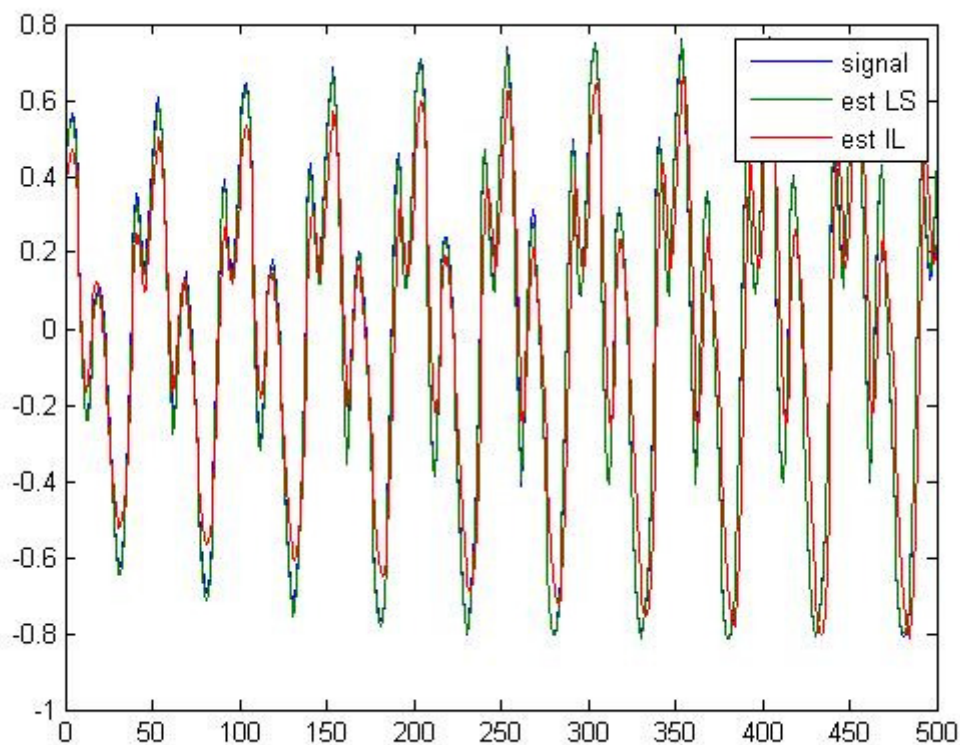


Se han realizado dos tipos de interpolación lineal:

- Una se muestra en rojo, obtiene por interpolación los valores de $(a_0 \text{ y } b_0)$ y a partir de estos valores $A = \sqrt{a_0^2 + b_0^2}$.
- Otra se muestra en verde, y obtiene directamente $A = \sqrt{a_0^2 + b_0^2}$ a partir de los valores de A de las tramas vecinas.

Se observa que el valor de la segunda interpolación (verde) es mucho más preciso, pero tiene el inconveniente de que los coeficientes del modelo (a_0 y b_0) permanecen desconocidos. Por lo tanto si queremos reconstruir la señal con los valores obtenidos mediante interpolación lineal, sólo podremos utilizar el primer tipo de interpolación (en rojo). El segundo tipo de interpolación sólo sería útil para utilizarla como restricción del sistema de ecuaciones (no como solución).

A continuación se muestra la reconstrucción de la señal utilizando los valores de la interpolación (en rojo), en la misma gráfica con la señal original (azul) y la estimación por mínimos cuadrados (verde):



4.4.5 Modificación de la base de tiempos del modelo

Hemos observado la gran variabilidad de los coeficientes de mayor orden (a_3 y b_3), que adquieren valores muy grandes perjudicando la precisión de la interpolación. Para un armónico concreto tenemos:

$$s(n) = (a_0 + a_1 n + a_2 n^2 + a_3 n^3) \sin 2\pi f_0 n + (b_0 + b_1 n + b_2 n^2 + b_3 n^3) \cos 2\pi f_0 n$$

donde n es la base de tiempos que adquiere los valores: $t = (0:N-1)$, donde N es el tamaño de la ventana de análisis.

Para intentar mitigar este efecto, hemos hecho varias modificaciones en la base de tiempos.

Una de las opciones que claramente produce que los coeficientes adquieran valores mucho más pequeños, es la modificación del modelo, que utiliza potencias positivas de n , por potencias negativas de n . Es decir:

$$s(n) = (a_0 + a_1 n^{-1} + a_2 n^{-2} + a_3 n^{-3}) \sin 2\pi f_0 n + (b_0 + b_1 n^{-1} + b_2 n^{-2} + b_3 n^{-3}) \cos 2\pi f_0 n$$

El inconveniente de esta modificación, es que la base de tiempos que antes era $t = (0:N-1)$, ahora no puede tomar el valor 0, porque genera indeterminación. Una solución posible es modificar la base de tiempos, por ejemplo dándole los valores:

$$t = (-N/2+0.5 : N/2-0.5)$$

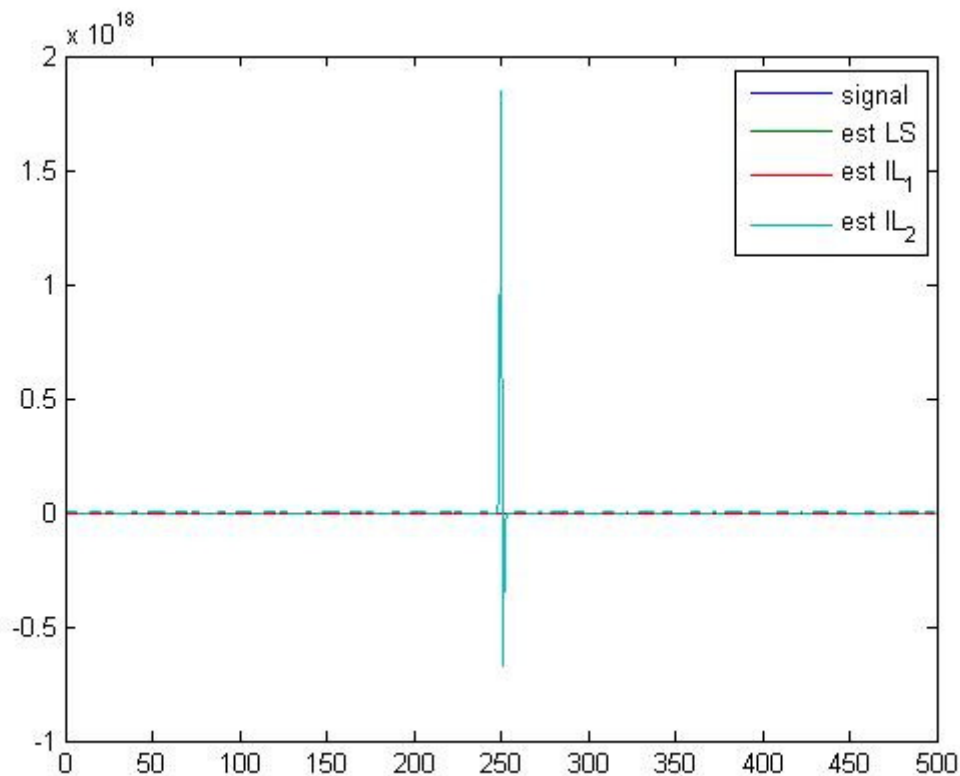
$$t = (N : 2N-1)$$

$$t = (2N : 3N-1)$$

...

Pero los resultados que se obtienen al realizar la estimación de la señal con estas modificaciones no son nada satisfactorios. Como se observa en la imagen siguiente, se producen inestabilidades gravísimas.

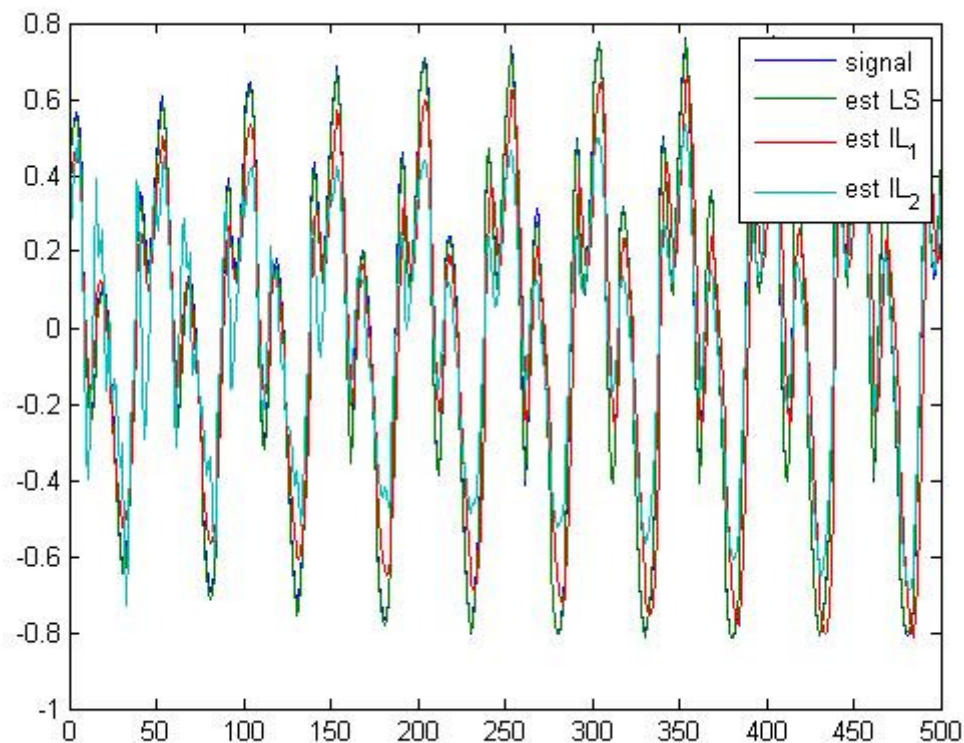
Estimación IL_2 : Modelo exponentes negativos, $t = (-N/2+0.5 : N/2-0.5)$



Por lo tanto, abandonando la idea de los exponentes negativos, las únicas modificaciones que podemos hacer en la base de tiempos son manteniendo el modelo de exponentes positivos.

En la siguiente gráfica, se muestra una trama de la señal original, y varias estimaciones utilizando coeficientes obtenidos por:

- IL_2 : interpolación lineal, con base de tiempos del modelo $t=(-N/2+0.5: N/2-0.5)$
- IL_1 : interpolación lineal, con base de tiempos del modelo $t=(0: N-1)$
- LS: mínimos cuadrados



En el apartado 6 analizaremos los resultados y valoraremos las distintas estimaciones realizando un estudio comparativo.

4.5 Filtro de Kalman

El segundo método que vamos a estudiar basa su funcionamiento en el filtro de Kalman. La descripción de las señales en este caso también hace uso del modelo sinusoidal, definiendo la señal para una trama $s(n)$ como la suma de las parciales más una componente residual:

$$s(n) = \sum_{i=1}^I A_i(n) \cos[2\pi i F_0(n)n + \varphi_i] + r(n), \quad n = 0, 1, \dots, N-1.$$

donde la parte determinista se corresponde con la superposición de I componentes armónicos (parciales), con amplitudes y frecuencias variantes en el tiempo. De la misma forma que en el apartado anterior, el algoritmo del filtro de Kalman también utiliza el modelado por aproximación polinomial. En este caso, la parte residual, se descompone en dos tipos de ruido:

- Ruido de medida: Debido al error de observación.
- Ruido proceso: Debido al error del modelo.

Como veremos más adelante, el buen funcionamiento de este algoritmo depende en gran medida de la correcta elección de los valores iniciales del ruido de medida y el ruido de proceso.

La característica fundamental de este método es que utiliza el filtro de Kalman para estimar y predecir los valores de los coeficientes del modelo para cada parcial. El filtro de Kalman es un algoritmo que toma la señal ruidosa observada (que en nuestro caso son las parciales identificadas), y tras evaluar la evolución de los estados posibles, realiza la estimación óptima del comportamiento del proceso. Se representa el modelo como:

$$\begin{aligned} x(k+1) &= Ax(k) + Bv(k) \\ y(k) &= Cx(k) + w(k) \end{aligned}$$

Donde

$$\begin{aligned} x(k) &= [f(k) \quad p(k) \quad n_1(k) \quad \dots \quad n_m(k)]^T \\ v(k) &= [u_1(k) \quad \dots \quad u_m(k)]^T \\ y(k) &= [f(k) \quad p(k)] \end{aligned}$$

Donde $f(k)$ y $p(k)$ son respectivamente la frecuencia y la potencia para una parcial detectada, $v(k)$ y $w(k)$ son el ruido de proceso y el ruido de medida, y $n_i(k)$ son el conjunto de estados para los cuales los ruidos de proceso $u_i(k)$ son blancos. A es la matriz de transición, B describe el acople del ruido de proceso $v(k)$ en estados del sistema, y C es la matriz de observación. En este modelo, $v(k)$ y $w(k)$ son procesos gaussianos incorrelados de media cero, y con matrices de covarianza Q y R .

El filtro de Kalman permite identificar el estado $x(k)$ a partir de las mediciones anteriores. Por lo tanto, habrá un periodo de enganche al principio de cada trama, hasta que se estabilice el filtro y consiga realizar una estimación más precisa.

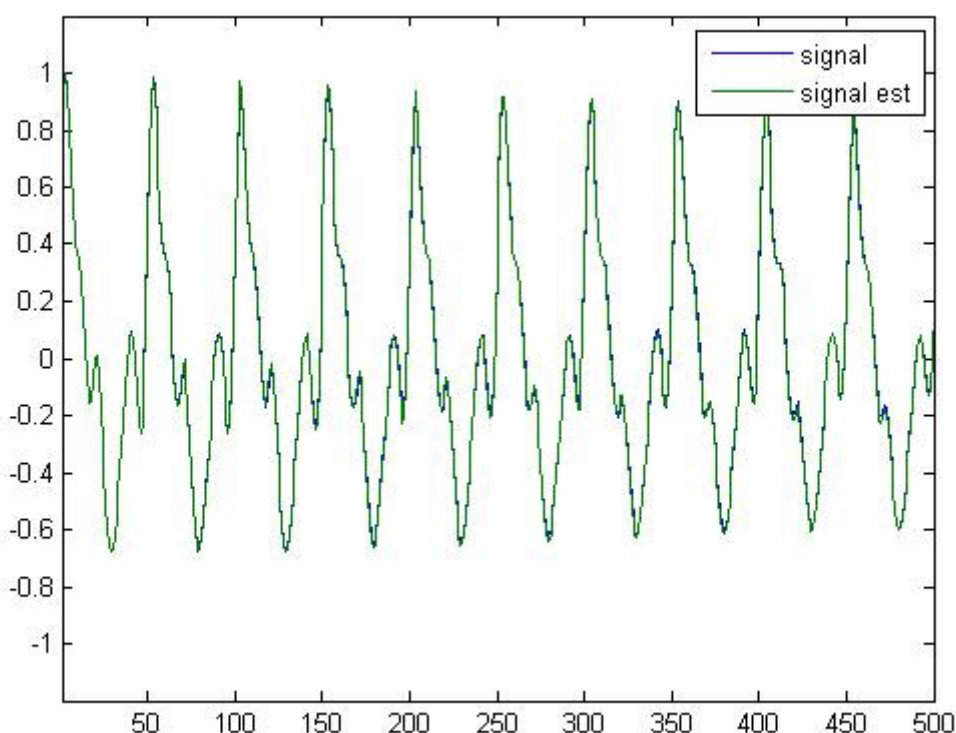
4.6 Resultados del filtro de Kalman

A continuación se muestra el comportamiento del filtro de Kalman, concretamente la estimación de una señal de flauta con tramas de 500 muestras. Se ha hecho un estudio con 10 armónicos.

Como hemos adelantado en el apartado anterior, el buen funcionamiento de este algoritmo depende en gran medida de la correcta elección de los valores iniciales del ruido de medida y el ruido de proceso. Tras realizar varias pruebas con distintos valores, hemos escogido:

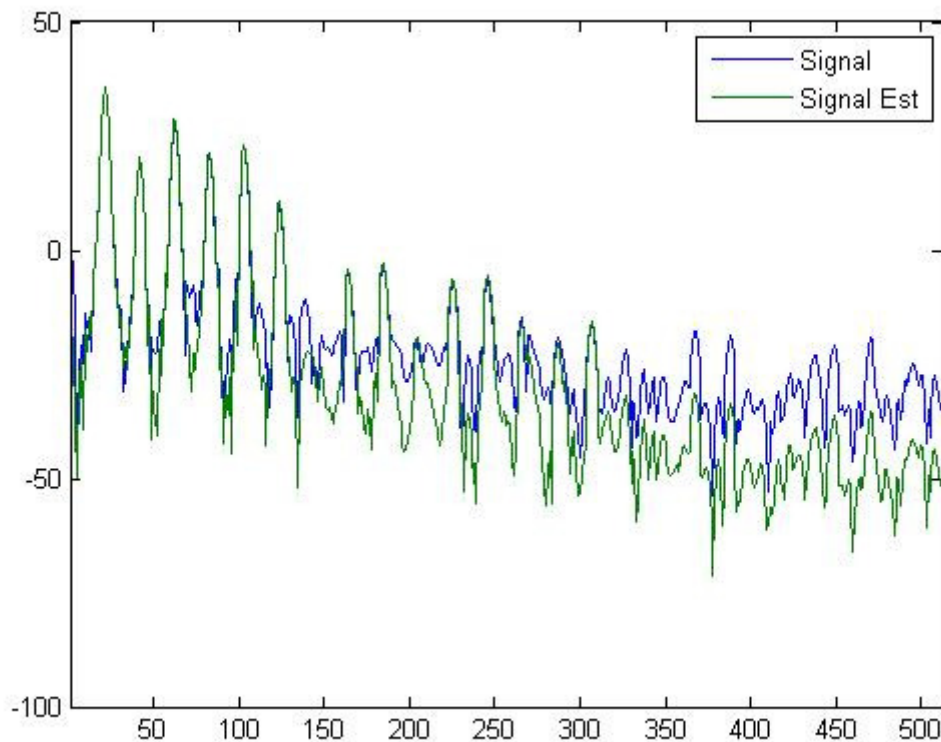
- SIGMA = 1 como parámetro inicial de la matriz de covarianza Q, que define el ruido de proceso.
- sigma = 0,1 como parámetro inicial del ruido de medida.

A continuación se muestra una trama de la señal original (azul) y su correspondiente estimación (verde) en dominio temporal.



Se observa la gran precisión con la que funciona este método. En el apartado 6 analizaremos los resultados y valoraremos las distintas estimaciones realizando un estudio comparativo.

Espectro de la señal anterior y su estimación:



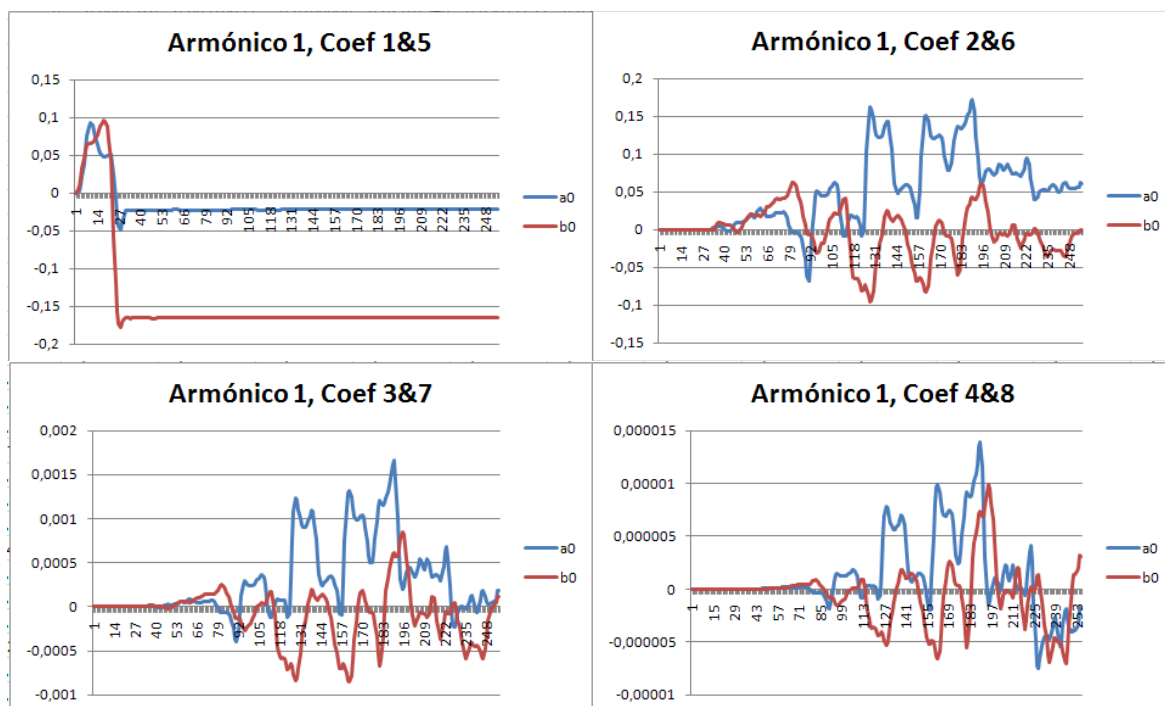
Se observa que a pesar de realizar un estudio con 10 armónicos, el espectro de la estimación tiene energía en frecuencias superiores a $10F_0$, a diferencia de lo que pasaba al utilizar el modelado polinomial visto anteriormente.

4.7. Estudio de los coeficientes del filtro de Kalman

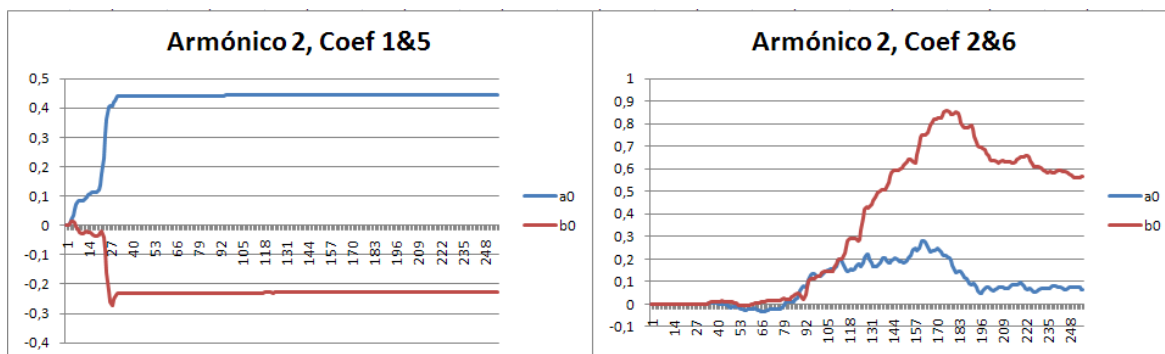
Para comprender mejor el funcionamiento del algoritmo, de manera similar al apartado anterior, llevaremos a cabo un estudio de la evolución temporal de los coeficientes del para una señal cuasi-estacionaria, pero esta vez realizaremos el estudio para una sola trama. En las gráficas siguientes se observa la evolución temporal de los coeficientes al avanzar las muestras dentro de una trama, y se muestran los coeficientes emparejados de la siguiente manera: (a_0, b_0) ; (a_1, b_1) ; (a_2, b_2) ; (a_3, b_3) ;

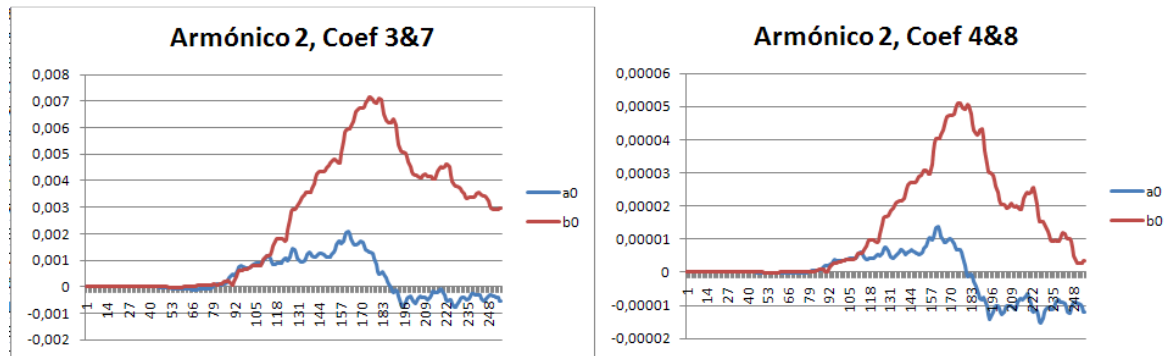
$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\sin 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\cos 2\pi f_0n$$

Armónico 1:

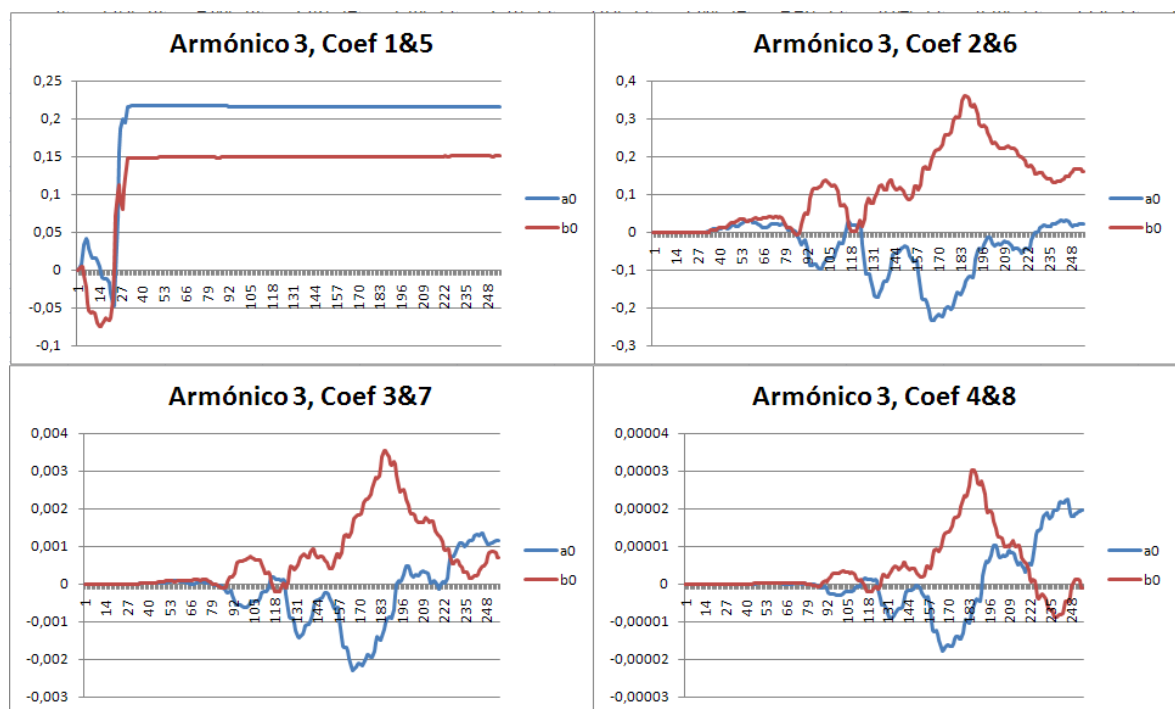


Armónico 2:





Armónico 3:



Si continuamos observando los coeficientes correspondientes al resto de armónicos, podemos ver que se comportan siguiendo el mismo patrón.

Se observa que la evolución temporal de los coeficientes dentro de una trama, siguen la misma forma para cada armónico aunque con distinta escala, exceptuando el primer coeficiente (a_0, b_0), que después del periodo de enganche en las primeras muestras, se estabiliza tomando un valor constante.

Hay que tener en cuenta también que es necesario introducir en el algoritmo los valores iniciales de los coeficientes del modelo, para realizar la estimación de la primera muestra de cada trama. Hay varias opciones para esto, una es darle el valor que tenían los coeficientes en la última muestra de la trama anterior, lo que beneficia la continuidad de la estimación. Pero nosotros le hemos dado el valor inicial cero a los coeficientes

correspondientes a la primera muestra de cada trama, porque vamos a hacer un análisis en el que no suponemos información previa conocida.

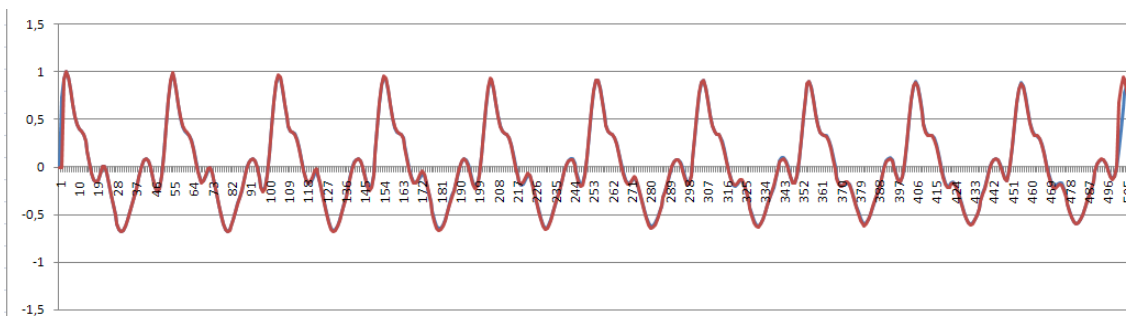
Al comienzo de cada trama, se observa que la estimación en las primeras muestras de la trama no es muy buena, pero va mejorando conforme avanzan las muestras. Este efecto se hace más patente en el estudio realizado en el apartado siguiente donde se realiza la separación de fuentes sonoras, y es debido al periodo de enganche del filtro de Kalman al comienzo de cada trama. Hasta que no se estabiliza el filtro no consigue realizar una estimación más precisa.

Se han estudiado varias opciones para mitigar este efecto:

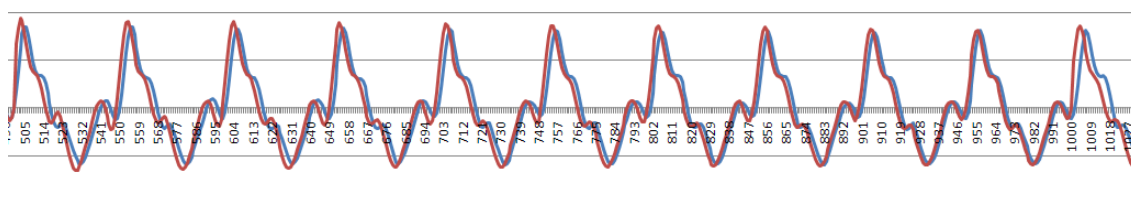
- Analizar la señal completa, sin segmentarla en tramas, y actualizando el valor de F0 cada 500 muestras

En las imágenes siguientes se muestra la señal original (azul) y su correspondiente estimación (rojo).

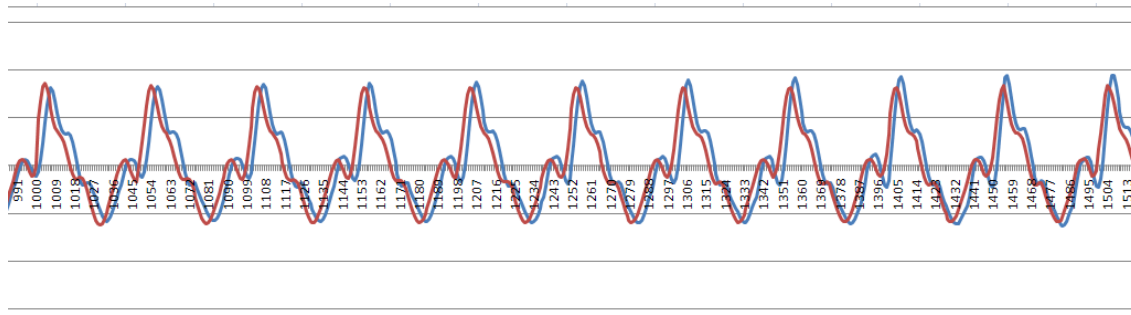
Muestras 1:500



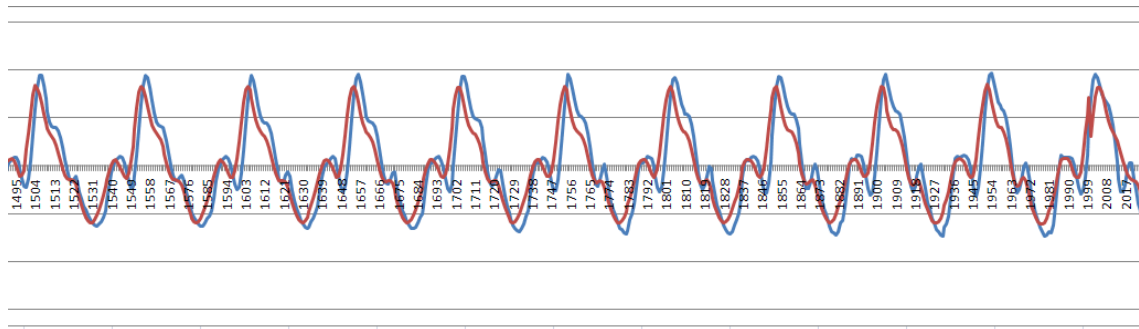
Muestras 500:1000



Muestras 1000:1500



Muestras 1500:2000



Este desfase entre la señal original y la estimada se va incrementando conforme avanzamos en el tiempo. Para intentar evitar esto se ha probado a no reinicializar los valores del ruido de proceso y ruido de medida, sino darles el valor que tenían en la última muestra de la trama anterior. Pero estas modificaciones no han dado buenos resultados, debido a que el funcionamiento del filtro de Kalman requiere un análisis por tramas donde se reinicialicen los valores de los coeficientes.

5 Separación de señales cuasi-armónicas procedentes de varias fuentes

5.1 Introducción

En este apartado se presenta el estudio realizado acerca de la estimación de señales cuasi-armónicas procedentes de dos fuentes. Por lo tanto, la señal $s(n)$ resultante se representa en el dominio temporal como la superposición de las señales $s_1(n)$ y $s_2(n)$, correspondientes a cada una de las fuentes que la componen:

$$s(n) = s_1(n) + s_2(n)$$

Sirviéndonos de los resultados obtenidos en la parte anterior, llevaremos a cabo la separación de fuentes sonoras con armónicos fuertemente solapados en frecuencia, mediante los siguientes métodos:

- Modelado polinomial
- Modelado polinomial con técnicas de interpolación
- Filtro de Kalman
- Filtrado en frecuencia

Al igual que en apartado anterior, se han utilizado señales aisladas de distintos instrumentos musicales obtenidas de una grabación en directo, y se ha realizado la mezcla de dichas señales normalizadas, de manera artificial. De esta manera tenemos la posibilidad de medir el grado de precisión de las señales estimadas respecto a las originales. De la misma forma que en el apartado anterior, las características de las señales utilizadas son:

- 44100 Hz de frecuencia de muestreo
- Codificación PCM de 8 bits por muestra
- 1 canal (señal mono)
- Ruido convolutivo del propio instrumento
- Ruido aditivo (de fondo) debido al ruido ambiente de la grabación

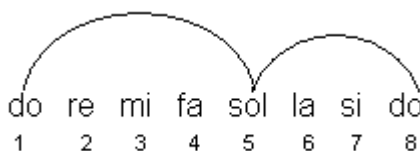
Además de estas características técnicas, supondremos que las señales utilizadas llevan asociadas modulaciones de amplitud y frecuencia (AM+FM). Como hemos mencionado en apartados anteriores, muchas situaciones se modelan de esta forma, como por ejemplo la presencia de ‘vibrato’ en la señal, que puede definirse como una AM inducida por la FM. Gracias a esto podremos representar las variaciones en amplitud, frecuencia y fase que sufren las señales, ya que a pesar de tratarse de señales cuasi-estacionarias, es inevitable la presencia de pequeñas variaciones temporales.

A continuación se estudiará el comportamiento de los distintos algoritmos mencionados, mostrando el resultado de la separación de dos fuentes sonoras:

- S1 = Señal de Trompeta nota C4 ($F_0 = 523,251$ Hz)
- S2 = Señal de Flauta nota G4 ($F_0 = 783,991$ Hz)

La razón de la elección de estas dos señales es que forman una quinta justa, o lo que es lo mismo, un intervalo musical compuesto por tres tonos y medio. Este tipo de intervalo lleva siendo estudiado desde Pitágoras, que fue quien le adjudicó un valor relativo de 3:2

en frecuencia. Es decir, para generar este intervalo deben sonar dos cuerdas de las mismas características cuyas longitudes tengan una relación de 1,5 o sonar dos tubos con una relación de alturas de 1,5. De la misma forma, las notas de que hemos escogido, cumplen aproximadamente la misma relación; aunque no exactamente, debido a razones de afinación y temperamento que no entraremos a discutir.



Si escuchamos una señal formada por una quinta, nos suena agradable precisamente por la relación de frecuencias que hay entre ambas, ya que comparten un gran número de armónicos. Cuanto más compleja sea esa relación, compartirán menos armónicos, y más disonantes sonarán los sonidos al producirse simultáneamente. Por esta razón, se considera la quinta como el intervalo más importante después de la octava (dos sonidos cuyas frecuencias fundamentales tienen una relación de 2:1).

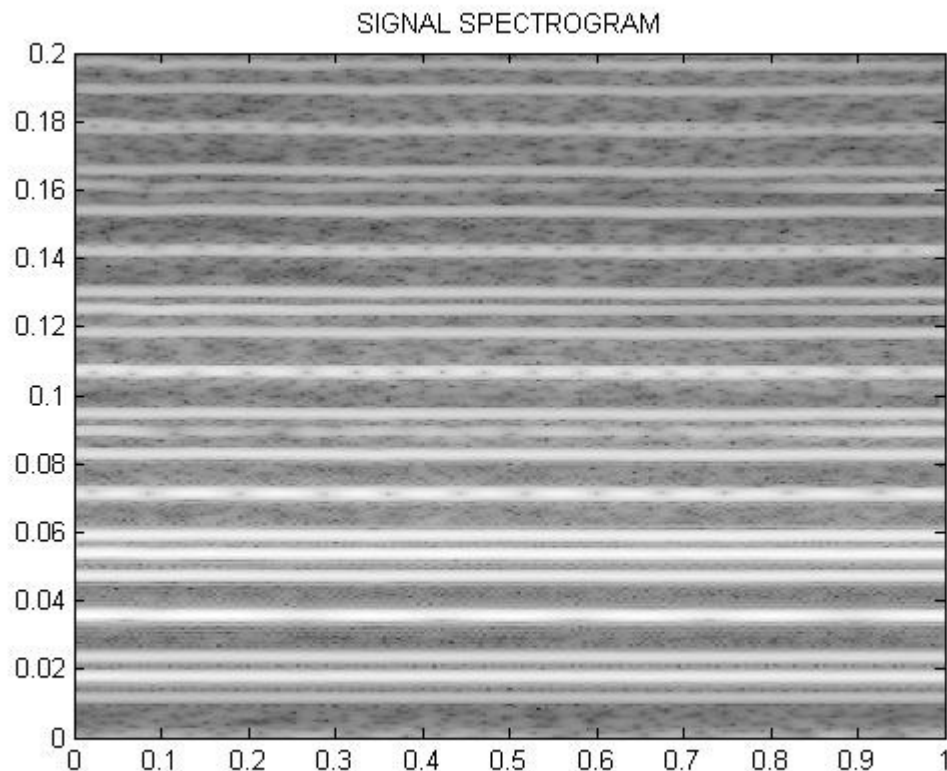
En la tabla siguiente se muestran los posibles intervalos que se pueden formar con una supuesta señal de 100 Hz, ordenados de forma que los que ocupan las posiciones superiores de la tabla comparten mayor número de armónicos. Los armónicos que se solapan aparecen en negrita:

		HARMÓNICOS									
Intervalo	Relación	1	2	3	4	5	6	7	8	9	10
Unísono	1:1	100.0	200.0	300.0	400.0	500.0	600.0	700.0	800.0	900.0	1000.0
Octava Justa	2:1	200.0	400.0	600.0	800.0	1000.0					
Quinta Justa	3:2	150.0	300.0	450.0	600.0	750.0	900.0				
Quarta Justa	4:3	133.3	266.6	400.0	533.3	666.6	800.0	933.3			
Sexta Mayor	5:3	166.6	333.2	500.0	666.6	833.3	1000.0				
Tercera Mayor	5:4	125.0	250.0	375.0	500.0	625.0	750.0	875.0	900.0		
Tercera menor	6:5	120.0	240.0	360.0	480.0	600.0	720.0	840.0	960.0		
Segunda Mayor	9:8	112.5	225.0	337.5	450.0	562.5	675.0	787.5	900.0		
Sexta menor	8:5	166.0	332.0	498.0	664.0	830.0	996.0				

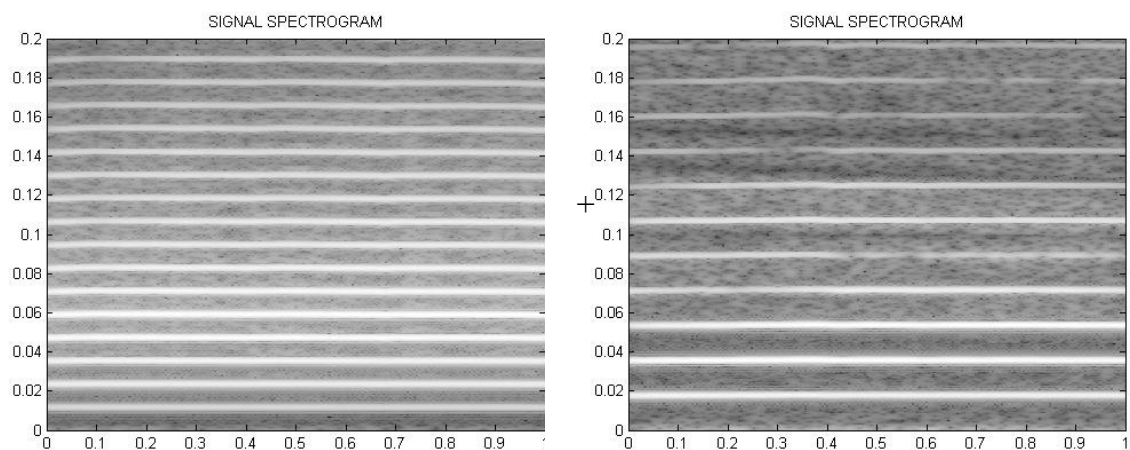
Debido a esto decimos que las señales escogidas son fuentes sonoras cuasi-armónicas. A continuación se muestra la señal de test que utilizaremos en los siguientes apartados, compuesta como hemos dicho por:

- Señal de Trompeta nota C4 ($F_0 = 523,251$ Hz)
- Señal de Flauta nota G4 ($F_0 = 783,991$ Hz)

Espectrograma de la señal de test compuesta por la mezcla (duración 1 segundo):

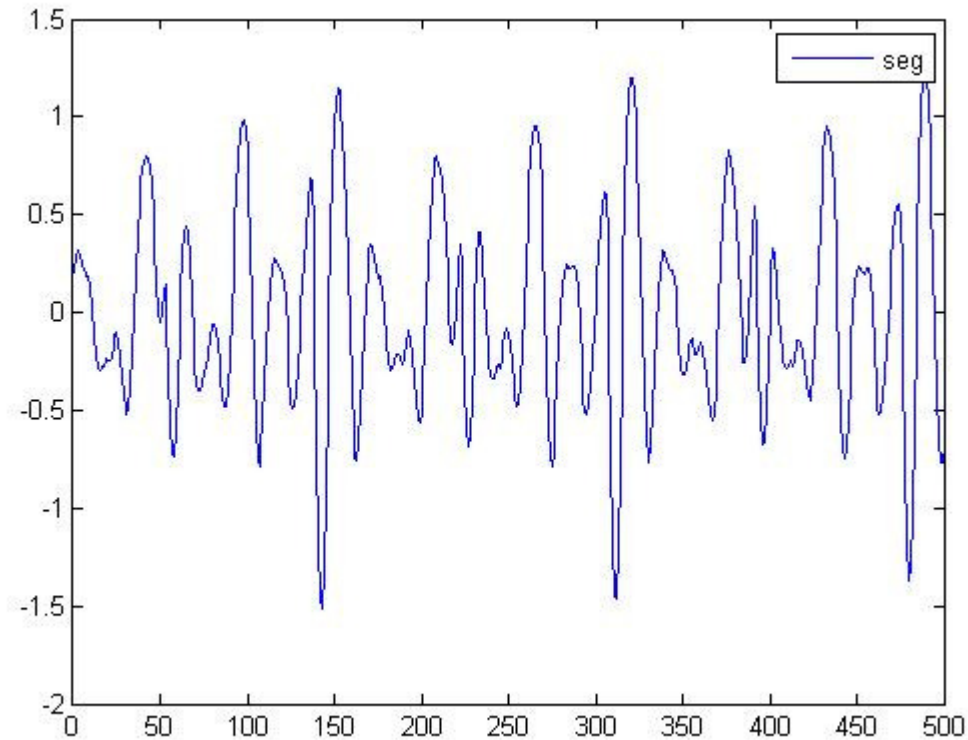


Espectrogramas de las señales que componen la señal de test anterior: señal de trompeta nota C4 (izquierda) + señal de flauta nota G4 (derecha).

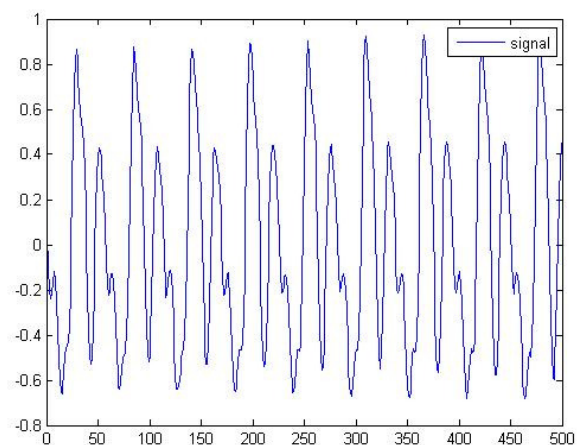
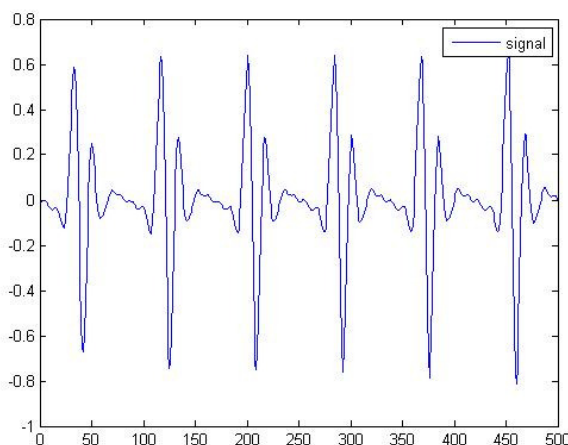


Se observa que los armónicos 2º, 4º y 6º de la señal de la derecha (G4) se solapan con los armónicos 3º, 6º y 9º de la señal de la izquierda (C4)

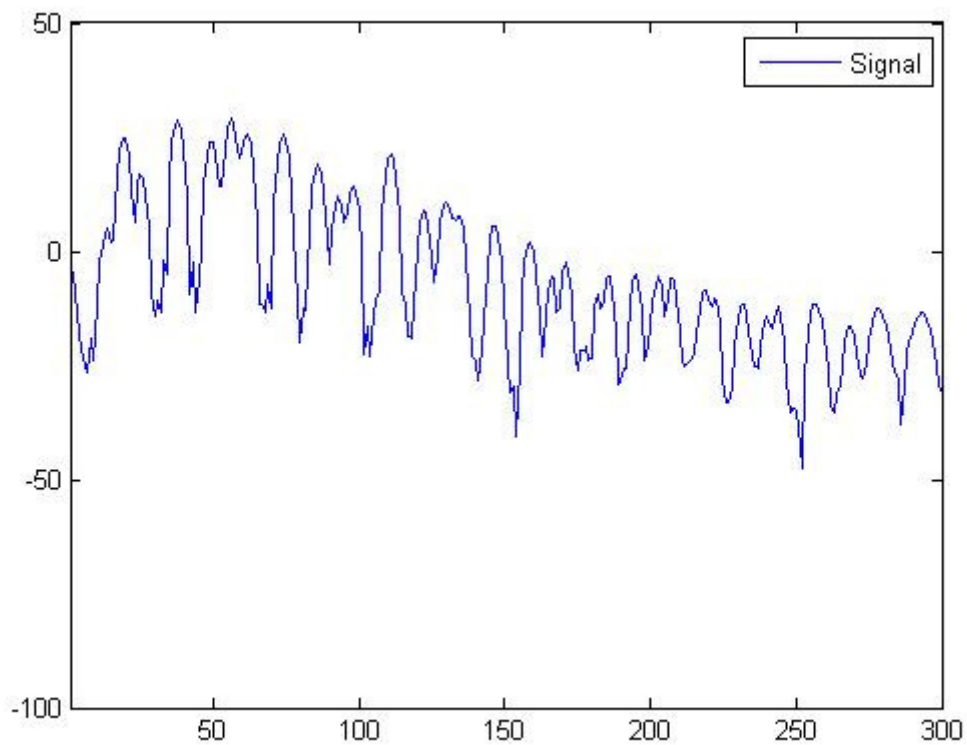
Señal de test, compuesta por la mezcla, en dominio temporal (una trama):



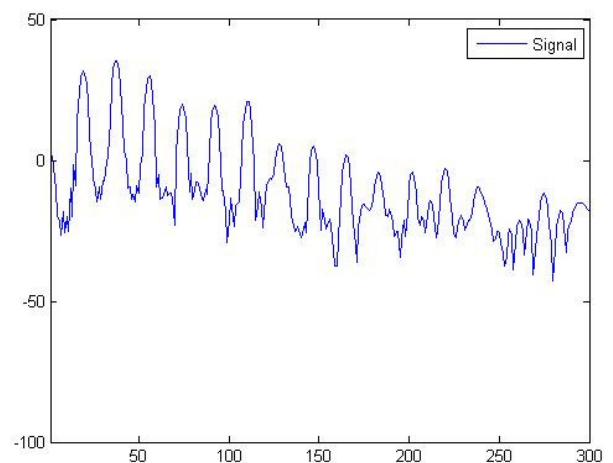
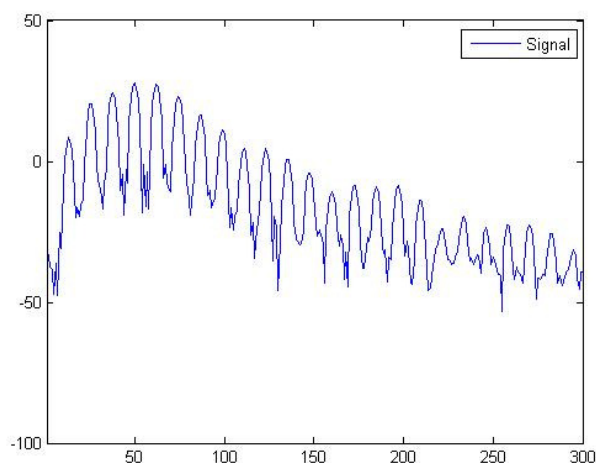
En la figura siguiente se muestran las señales S1, a la izquierda (Señal de Trompeta nota C4 con $F_0 = 523,251$ Hz) y S2, a la derecha (Señal de Flauta nota G4 con $F_0 = 783,991$ Hz), que componen la señal anterior:



Espectro de la señal de test (una trama):



Espectros de las señales S1, a la izquierda (Señal de Trompeta nota C4 con $F_0 = 523,251$ Hz) y S2, a la derecha (Señal de Flauta nota G4 con $F_0 = 783,991$ Hz), que componen la señal anterior:



5.2 Modelado polinomial

En este apartado se presenta el estudio realizado acerca de la estimación de señales cuasi-armónicas procedentes de dos fuentes utilizando el modelado polinomial. El funcionamiento del algoritmo de separación es muy similar al visto anteriormente para señales simples, con la diferencia de que es necesario estimar dos frecuencias fundamentales F_0 , una para cada señal. Una vez estimada la frecuencia fundamental de cada señal se procede de la misma manera que en el tratamiento de señales simples, obteniendo para cada señal los coeficientes del modelo mediante mínimos cuadrados (LS).

Del primer paso depende enormemente el grado de éxito del algoritmo, y se trata de realizar una buena estimación de las frecuencias fundamentales de cada señal. Hay varios métodos para realizar esta tarea, y dependiendo de las circunstancias particulares será aconsejable utilizar unos u otros. Para las condiciones en las que nos encontramos hemos utilizado un algoritmo de detección de F_0 basado en el método ‘ASA’, explicado anteriormente en el apartado 4.2, indicado para señales cuasi-estacionarias. Las frecuencias fundamentales obtenidas son:

- Señal de Trompeta nota C4: $F_0 = 521,9$ Hz.
- Señal de Flauta nota G4: $F_0 = 786,5$ Hz.

Como es lógico, hay una pequeña variación respecto a los valores teóricos:

- Señal de Trompeta nota C4 ($F_0 = 523,251$ Hz)
- Señal de Flauta nota G4 ($F_0 = 783,991$ Hz)

Esto no es debido a una mala estimación de F_0 , sino a que se trata de señales grabadas en directo de instrumentos musicales reales, con afinación imperfecta.

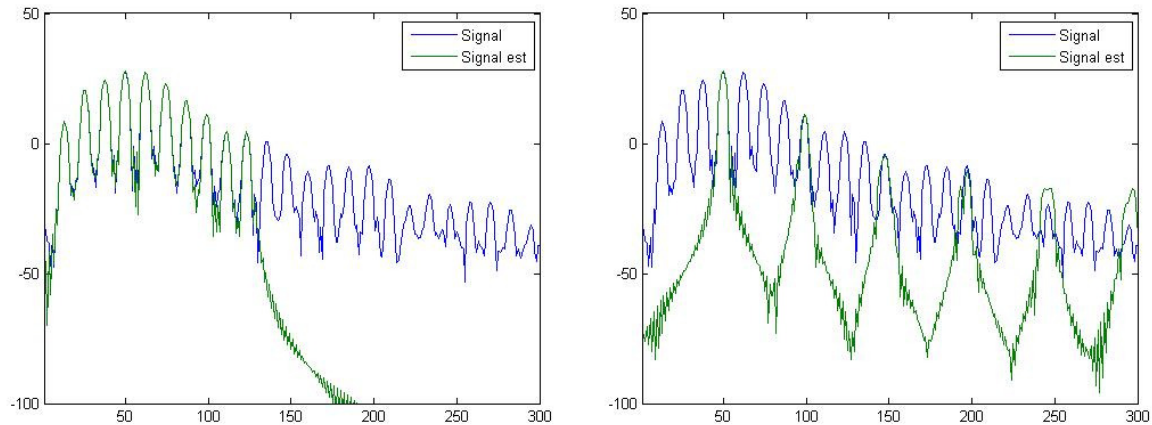
Si habríamos elegido otro algoritmo de detección de F_0 , es posible que hubiéramos obtenido otros valores diferentes, equivocándonos en la estimación. Por ejemplo, utilizando el algoritmo desarrollado por T. Grandke "*Interpolation algorithms for discrete Fourier transform of weighted signals*", obtenemos estos otros valores:

- Señal de Trompeta nota C4: $F_0 = 2607,6$ Hz ($= 5 \times 523,251$)
- Señal de Flauta nota G4: $F_0 = 2363,4$ Hz. ($= 3 \times 783,991$)

Queda claro que en estas circunstancias el método ‘Grandke’ es mucho menos preciso, pero hay que tener en cuenta que el método ‘ASA’ suele fallar o bien en ausencia de energía de la señal (pausas entre notas) o bien si el segmento analizado es altamente no estacionario. En cambio, el método ‘Grandke’ no tiene problemas de convergencia. Por lo tanto, cada método puede tener ventajas dependiendo de la situación en la que nos encontremos.

De todas formas, detectando mal la frecuencia fundamental de las señales no conseguiremos hacer una buena estimación, como se muestra en las imágenes siguientes:

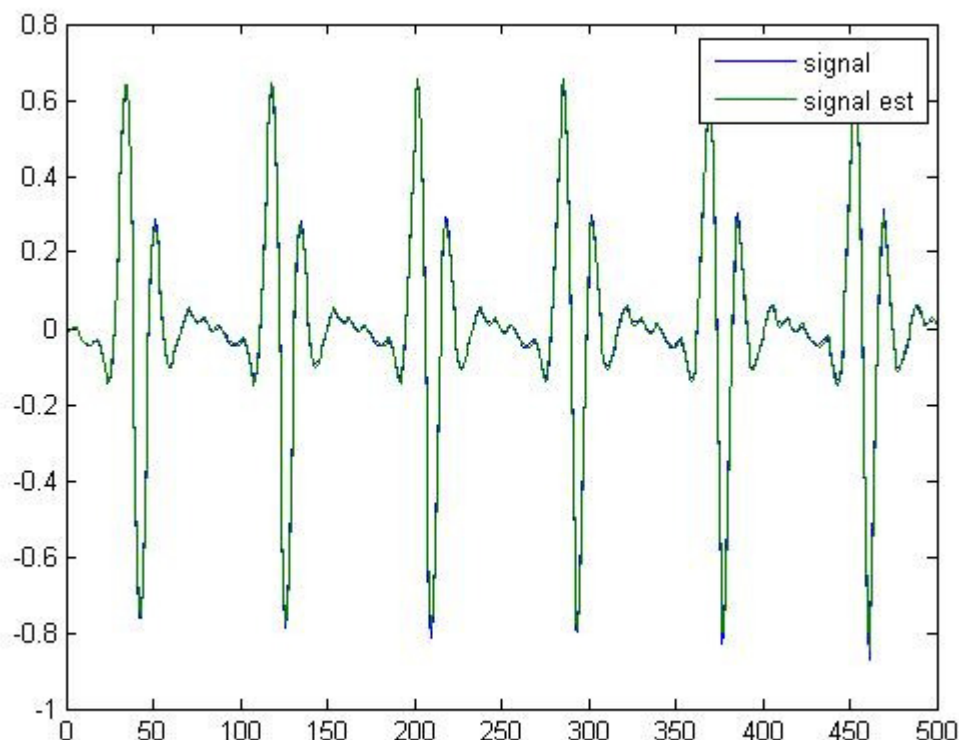
Espectro de la señal de trompeta nota C4 (azul) y su correspondiente estimación (verde) utilizando el modelado polinomial con el método de detección de frecuencia fundamental ‘ASA’ (izquierda), y utilizando el método ‘Grandke’ (derecha):



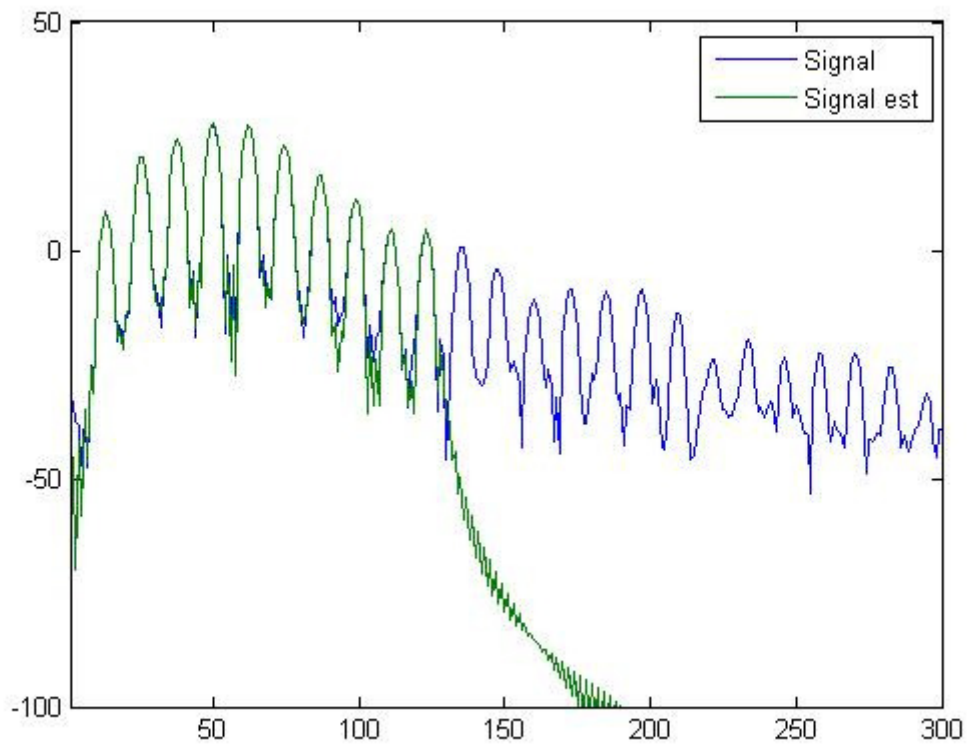
Como hemos visto en el apartado 4.2, una vez que tenemos el valor de la frecuencia fundamental F_0 , el resto de parámetros del modelo se obtienen mediante el método de mínimos cuadrados (LS).

A continuación se muestran los resultados tras realizar la separación de señales utilizando el método de modelado polinomial. En la siguiente imagen vemos el resultado de la estimación de las señales, comparadas con las señales originales. Se han utilizado tramas de 500 muestras, polinomios de orden 4, y se ha hecho un estudio con 10 armónicos.

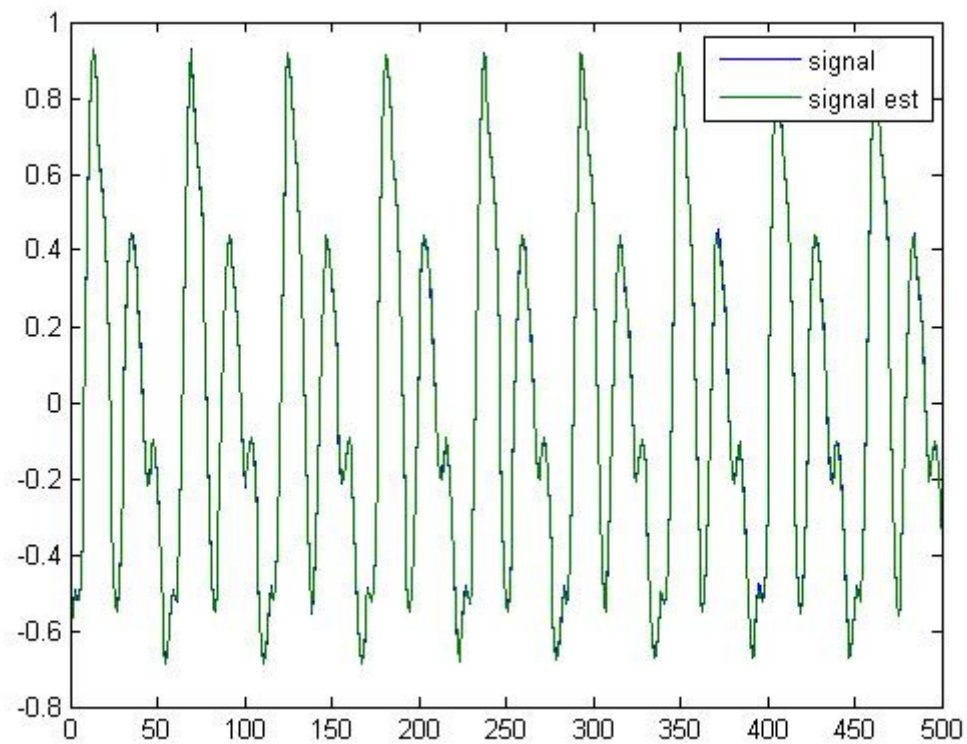
Señal de trompeta nota C4:



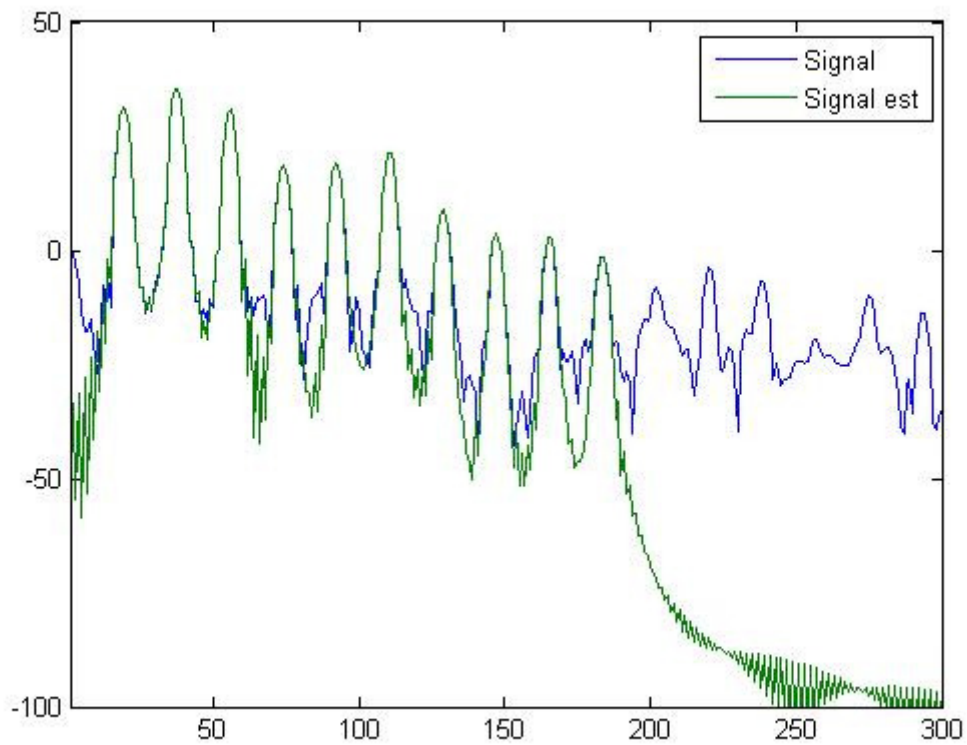
Espectro señal de Trompeta nota C4:



Señal flauta nota G4:



Espectro señal de flauta nota C4:



En el apartado 6 analizaremos los resultados y valoraremos las distintas estimaciones realizando un estudio comparativo.

5.3 Modelado polinomial con técnicas de interpolación

En este apartado hacemos uso de las técnicas de interpolación estudiadas en el apartado 4.4 para aplicarlas a la separación de fuentes sonoras cuasi-armónicas. El procedimiento es similar al utilizado para señales simples. Suponemos que la situación en la que nos encontramos es similar a la anterior: tenemos una señal de flauta y otra de trompeta separadas por un intervalo de quinta justa. Como hemos dicho anteriormente, en estas condiciones, las dos fuentes sonoras comparten armónicos que se solapan en frecuencia. Ante estas circunstancias, tanto el modelado polinomial como otros métodos suelen dar resultados poco satisfactorios. Además, este fenómeno es muy común en señales musicales, por lo que resulta necesario resolver este problema.

En estas circunstancias de solapamiento de armónicos, los parámetros del modelo polinomial obtenidos por mínimos cuadrados (LS) son erróneos, pero podemos aprovecharnos del criterio de continuidad temporal de las señales musicales para realizar una aproximación de dichos parámetros. Debido a que las señales musicales van variando suavemente en el tiempo cambiando de frecuencia y amplitud, puede darse la situación de que el solapamiento de armónicos se produzca solo en un intervalo de tiempo determinado, mientras que antes y después de dicho intervalo no existe solapamiento de armónicos. En estas circunstancias resulta claro que una buena aproximación de los parámetros del modelo podría realizarse mediante la interpolación de los valores de los instantes de tiempo anterior y posterior, donde los parámetros son correctos porque no hay solapamiento de armónicos.

Para simular esta situación procedemos de manera similar que en el apartado 4.4 donde realizábamos la estimación de los parámetros de una trama mediante interpolación lineal, a partir de los parámetros de las tramas adyacentes. Se han realizado dos tipos de interpolación, los estudiados en el apartado 4.4, con distintas bases de tiempos:

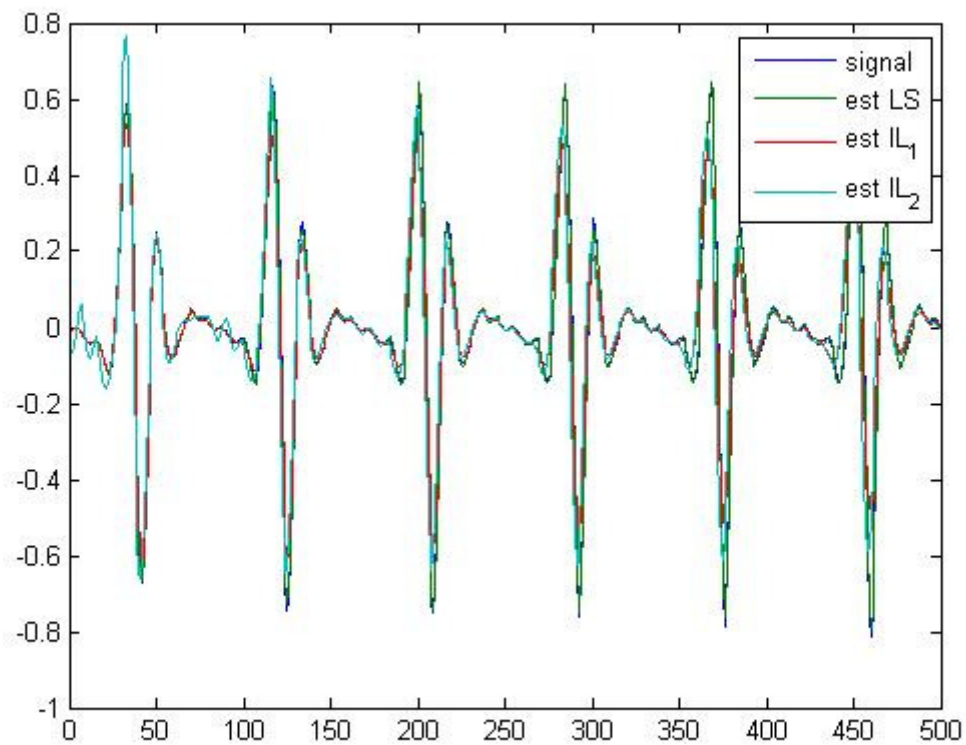
- IL_2 : interpolación lineal, con base de tiempos del modelo $t=(-N/2+0.5: N/2-0.5)$
- IL_1 : interpolación lineal, con base de tiempos del modelo $t=(0: N-1)$

Además de las estimaciones obtenidas por interpolación, en las gráficas siguientes se muestran la señal original y la estimación obtenida por mínimos cuadrados (LS).

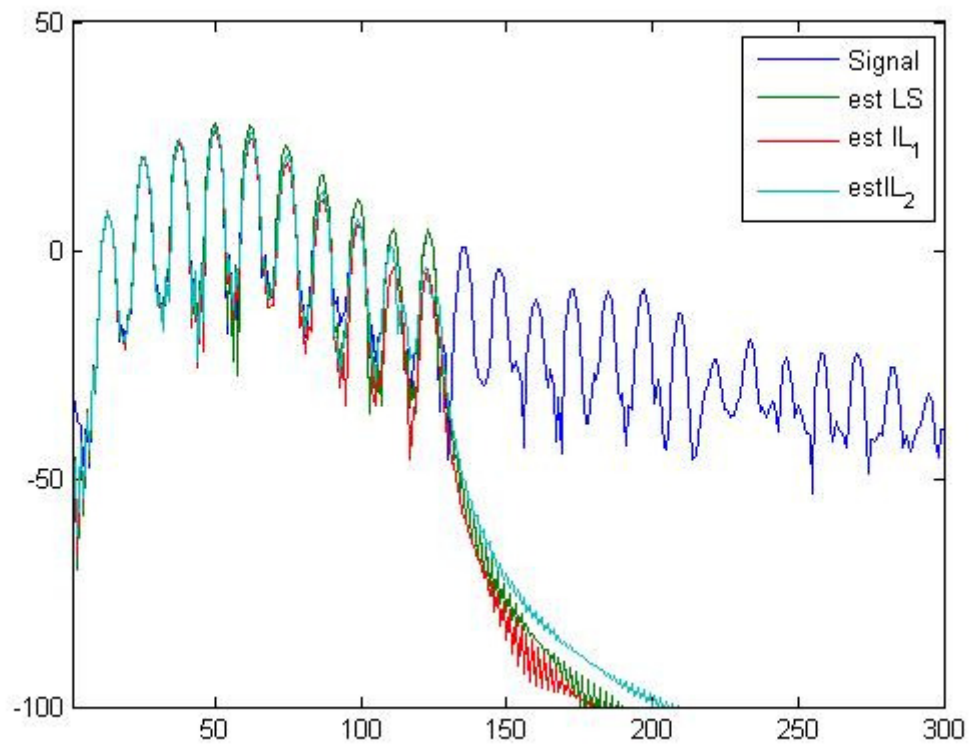
A continuación se muestra el resultado de la separación de dos fuentes sonoras:

- Señal de Trompeta nota C4 ($F_0 = 523,251$ Hz)
- Señal de Flauta nota G4 ($F_0 = 783,991$ Hz)

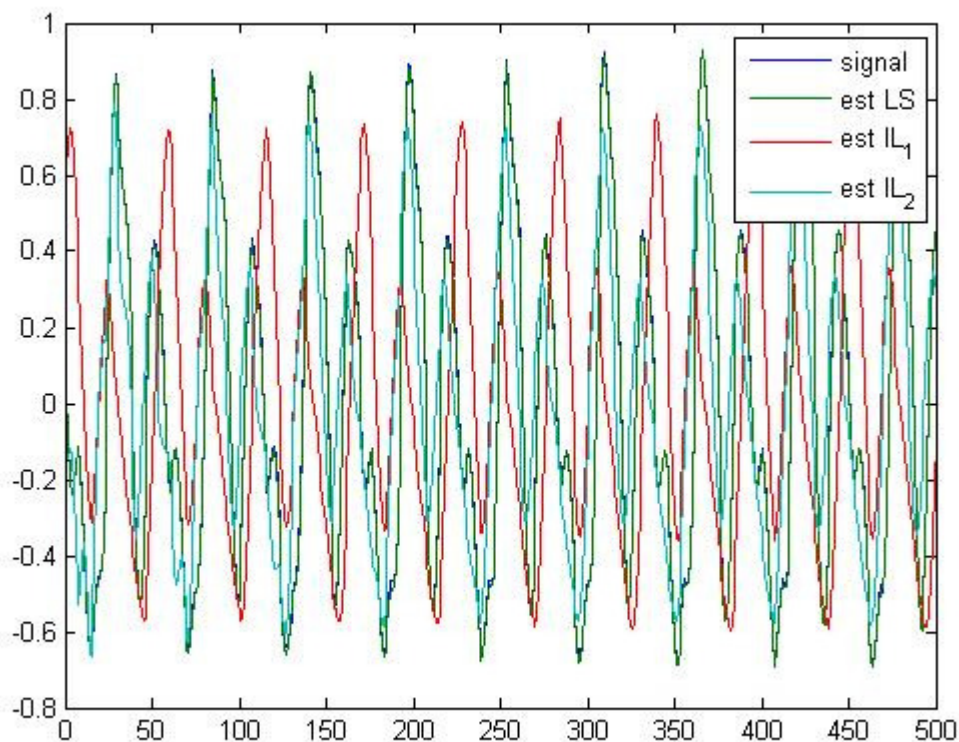
Señal de trompeta nota C4:



Espectro señal trompeta nota C4:

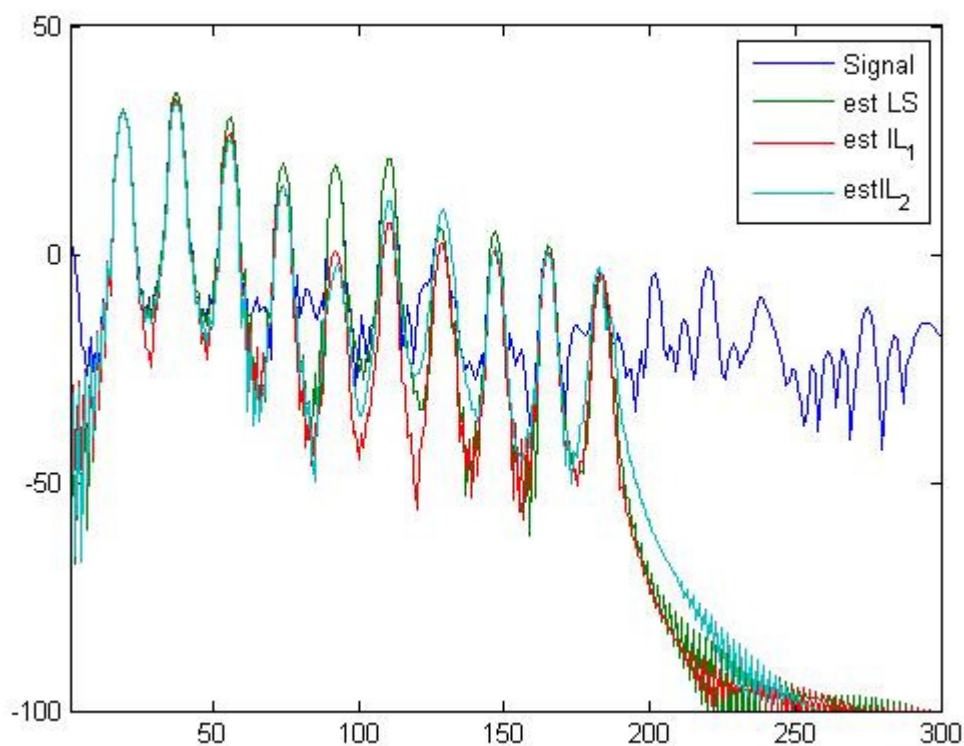


Señal flauta nota G4:



Observamos en la imagen anterior que la estimación por el método IL_2 (rojo) está desfasada respecto a la original.

Espectro señal flauta nota G4:



En el apartado 6 analizaremos los resultados y valoraremos las distintas estimaciones realizando un estudio comparativo.

5.4 Filtro de Kalman

En este apartado estudiaremos el comportamiento del método visto en el apartado 4.5, aplicado a la separación de fuentes sonoras cuasi-armónicas. El procedimiento es análogo al desarrollado para fuentes simples, con la diferencia de que para cada trama hay que estimar dos frecuencias fundamentales F_0 , una para cada señal. Una vez que tenemos las frecuencias fundamentales, se obtienen el resto de parámetros del modelo y se realiza la estimación de señales mediante el algoritmo del filtro de Kalman. Es decir, que utiliza el filtro de Kalman para estimar y predecir los valores de la amplitud de cada parcial en cada instante de tiempo.

Al igual que en el caso de tratar señales simples, el buen funcionamiento de este algoritmo depende en gran medida de la correcta elección de los valores iniciales del ruido de medida y el ruido de proceso. Tras realizar varias pruebas con distintos valores, hemos escogido:

- SIGMA = 1 como parámetro inicial de la matriz de covarianza Q, que define el ruido de proceso.
- sigma = 0,5 como parámetro inicial del ruido de medida.

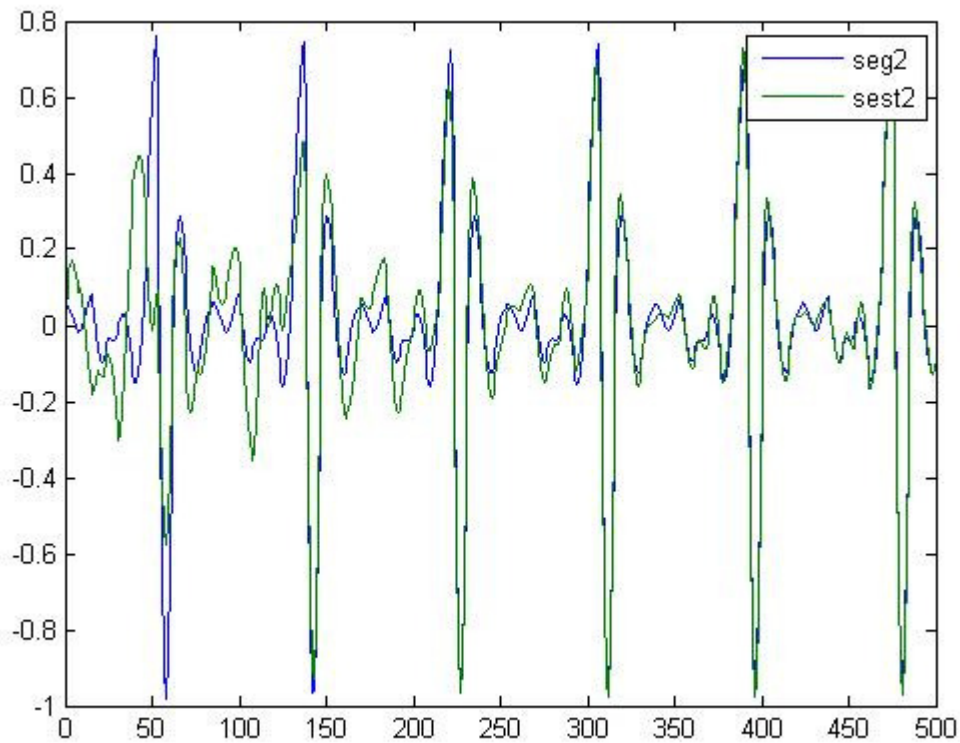
También hay que introducir en el algoritmo los valores iniciales de los coeficientes del modelo, para realizar la estimación de la primera muestra de cada trama. Hay varias opciones para esto, una es darle el valor que tenían los coeficientes en la última muestra de la trama anterior, lo que beneficia la continuidad de la estimación. Pero nosotros le hemos dado el valor inicial cero a los coeficientes correspondientes a la primera muestra de cada trama, porque vamos a hacer un análisis de una trama en el que no suponemos información previa conocida.

Para tratar de comparar los distintos métodos estudiados en este proyecto, seguiremos analizando la señal compuesta por una quinta justa formada por las señales:

- Señal de Trompeta nota C4 ($F_0 = 523,251$ Hz)
- Señal de Flauta nota G4 ($F_0 = 783,991$ Hz)

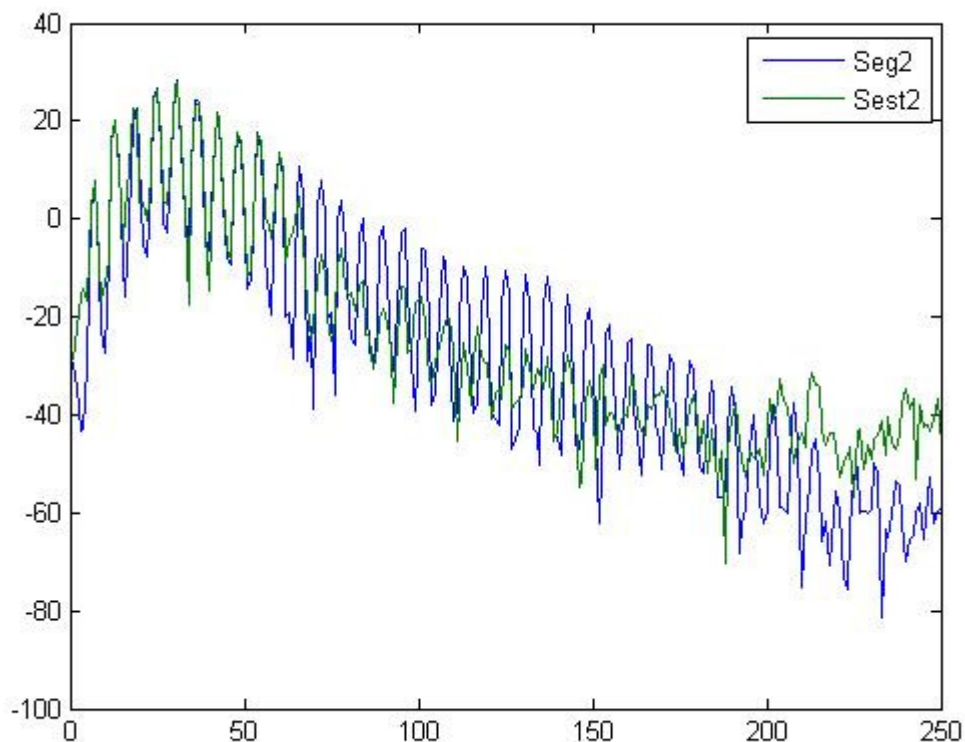
A continuación se muestra una trama de las señales originales (azul) y sus correspondientes estimaciones (verde) en dominio temporal. Se ha hecho un análisis con 10 armónicos.

Señal de trompeta nota C4:

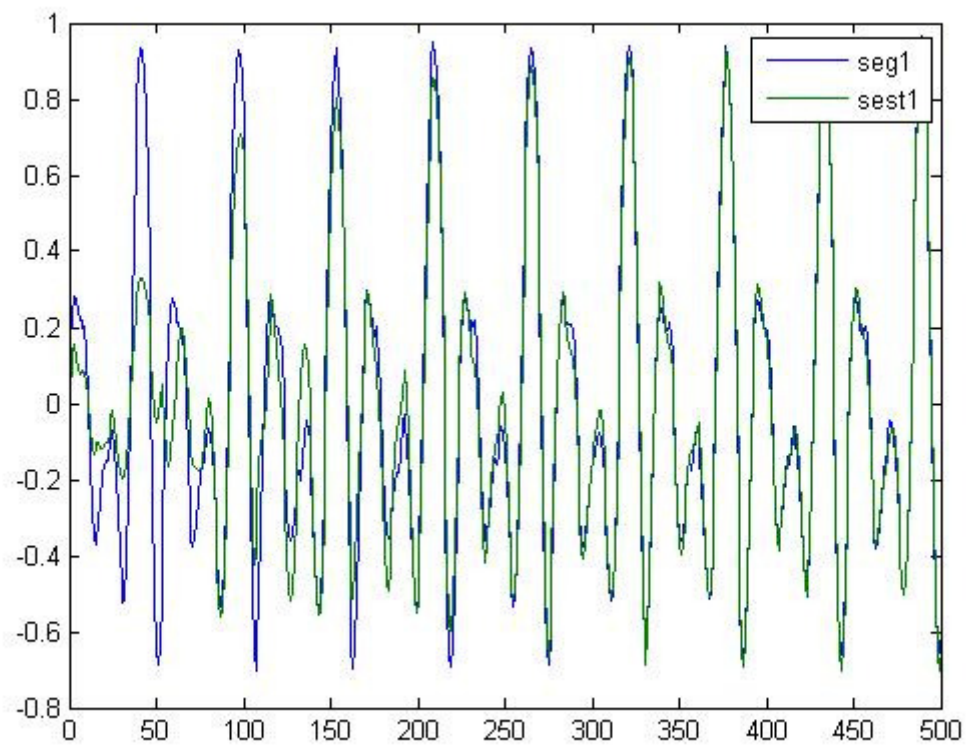


Se observa que la estimación en las primeras muestras de la trama no es muy buena, pero va mejorando conforme avanzan las muestras. Esto es debido al periodo de enganche del filtro de Kalman, y sucede al principio de cada trama, hasta que se estabiliza el filtro y consigue realizar una estimación más precisa.

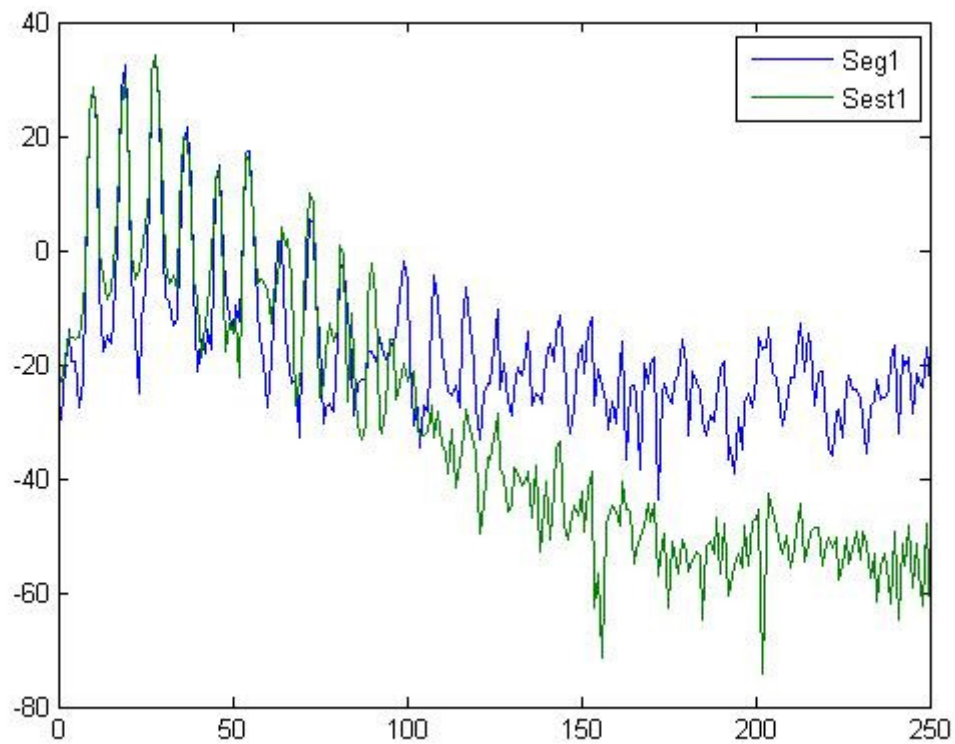
Espectro señal trompeta C4:



Señal flauta nota G4:



Espectro señal flauta nota G4

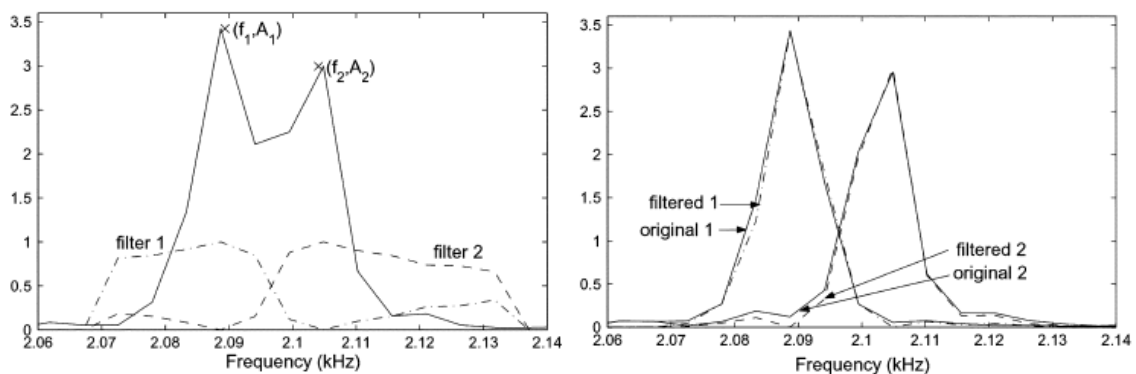


En el apartado 6 analizaremos los resultados y valoraremos las distintas estimaciones realizando un estudio comparativo.

5.5 Filtrado en frecuencia

En este apartado estudiaremos un nuevo método consistente en el filtrado en frecuencia, desarrollado originalmente por Mark R. Every y John E. Szymanski [22], y que fue publicado por el IEEE en septiembre de 2006.

El procedimiento que sigue este algoritmo consiste básicamente en filtrar el conjunto de armónicos que se solapan en cada trama, dividiendo el pico espectral donde se produce el solapamiento para asignar cada parte a su fuente correspondiente, como se muestra en la figura siguiente:

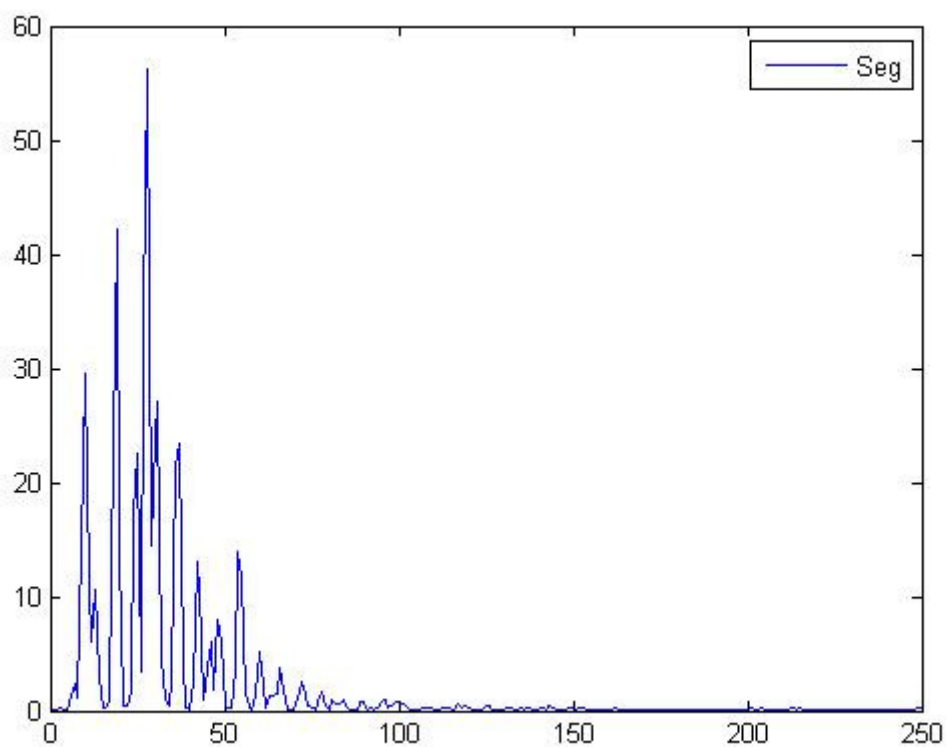


De la misma manera que en apartados anteriores, estudiaremos la señal compuesta por una quinta justa formada por las señales:

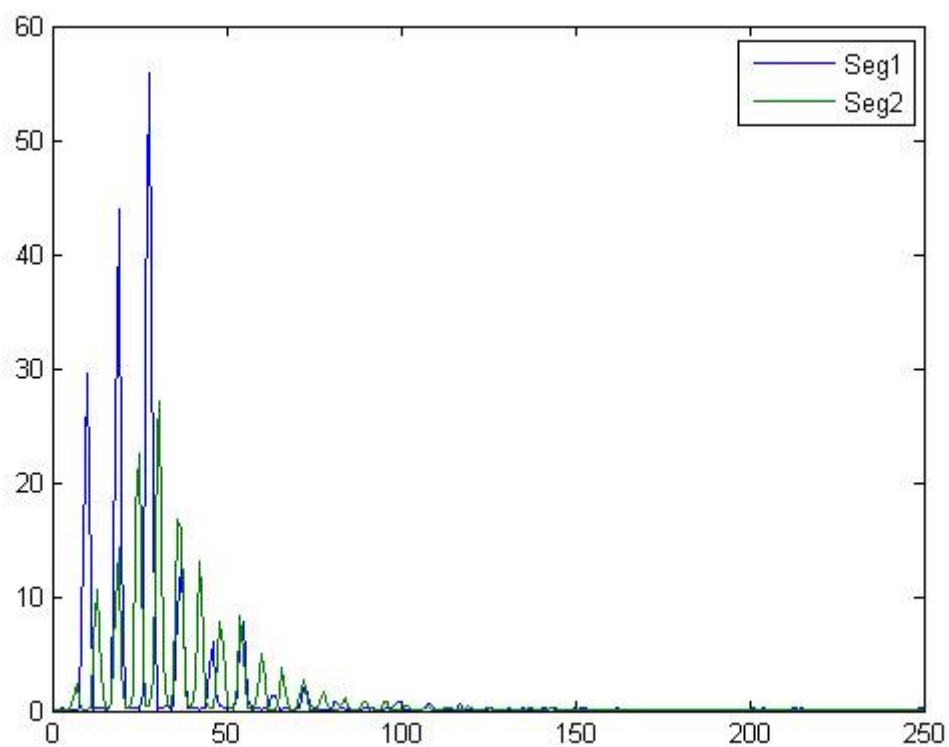
- Señal de Trompeta nota C4 ($F_0 = 523,251$ Hz)
- Señal de Flauta nota G4 ($F_0 = 783,991$ Hz)

A continuación se muestra una trama de las señales originales y sus correspondientes estimaciones:

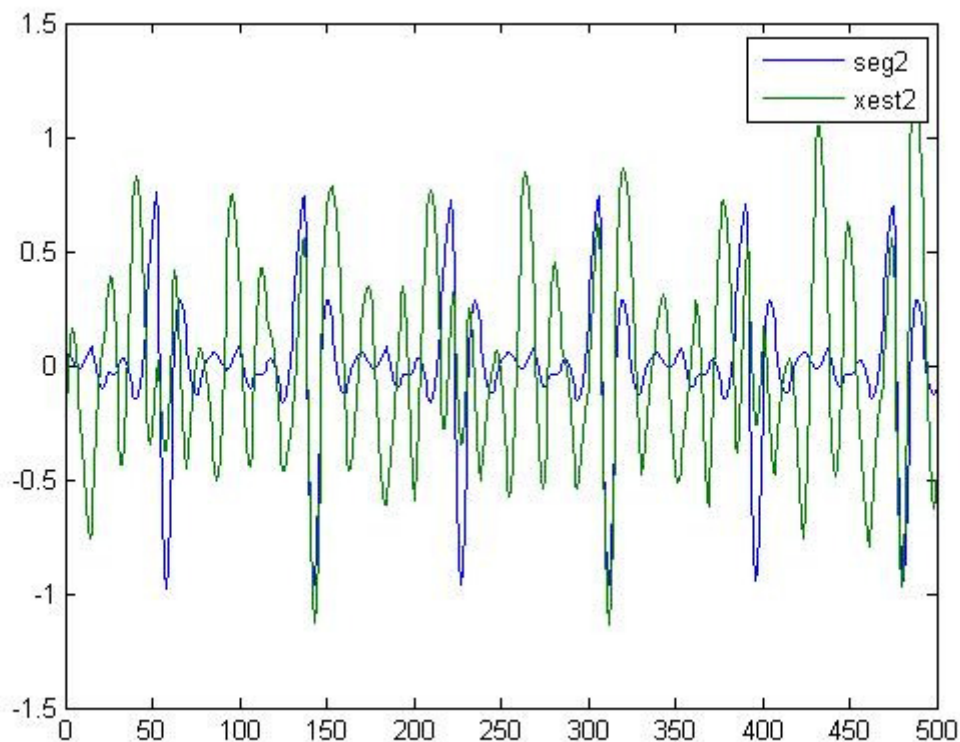
Espectro de la señal original compuesta por la mezcla de señales (una trama):



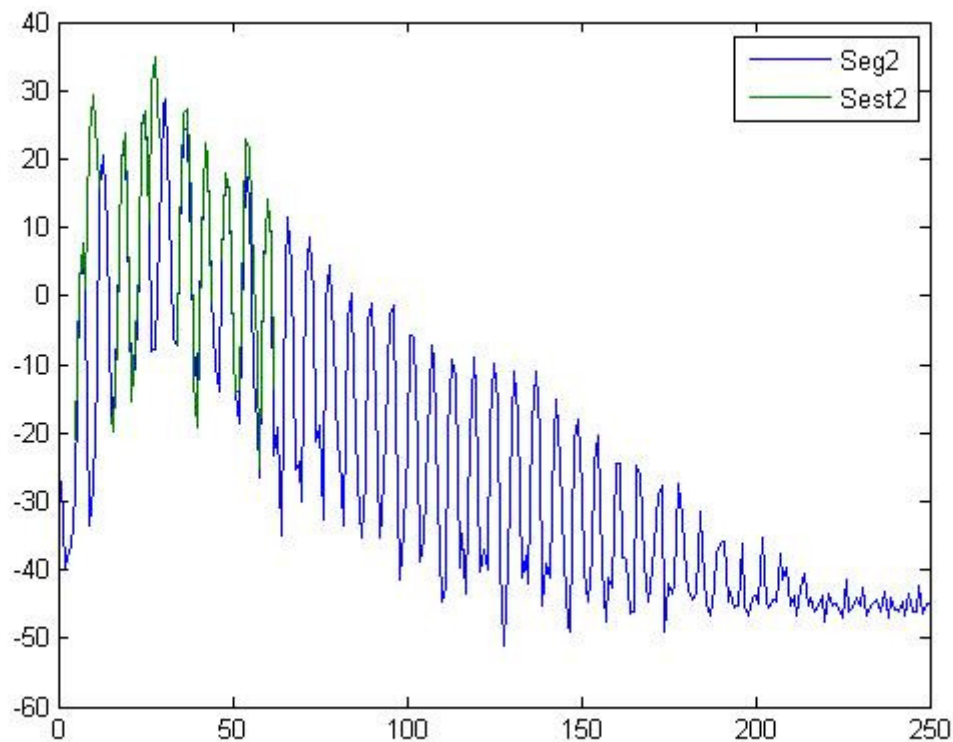
Descomposición del espectro anterior separando cada parte del espectro a su fuente correspondiente:



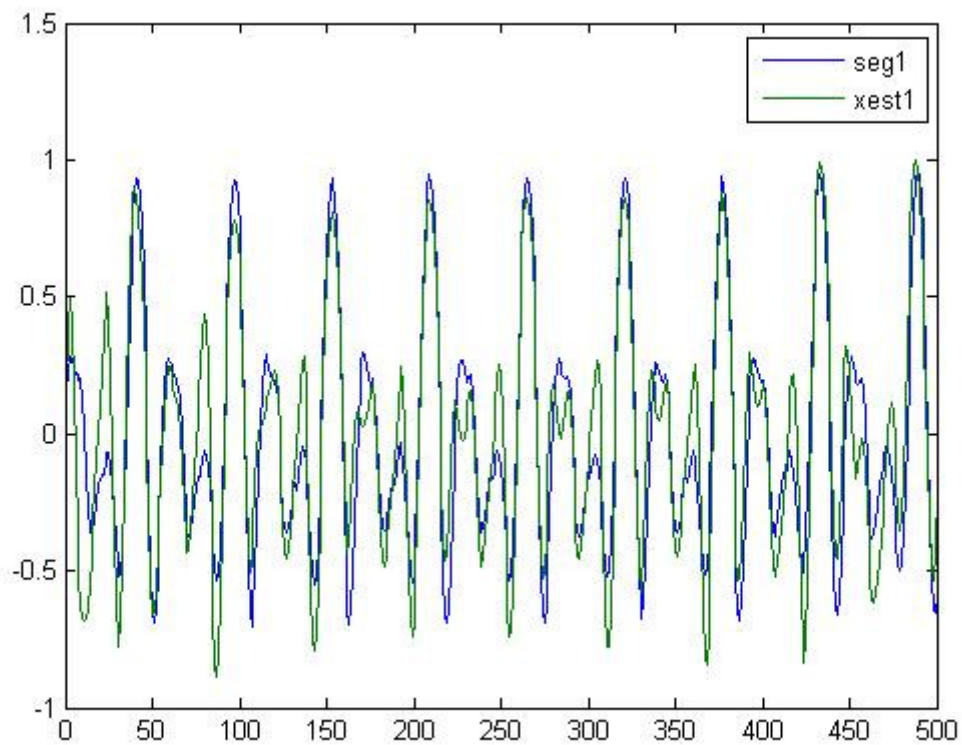
Señal de trompeta nota C4:



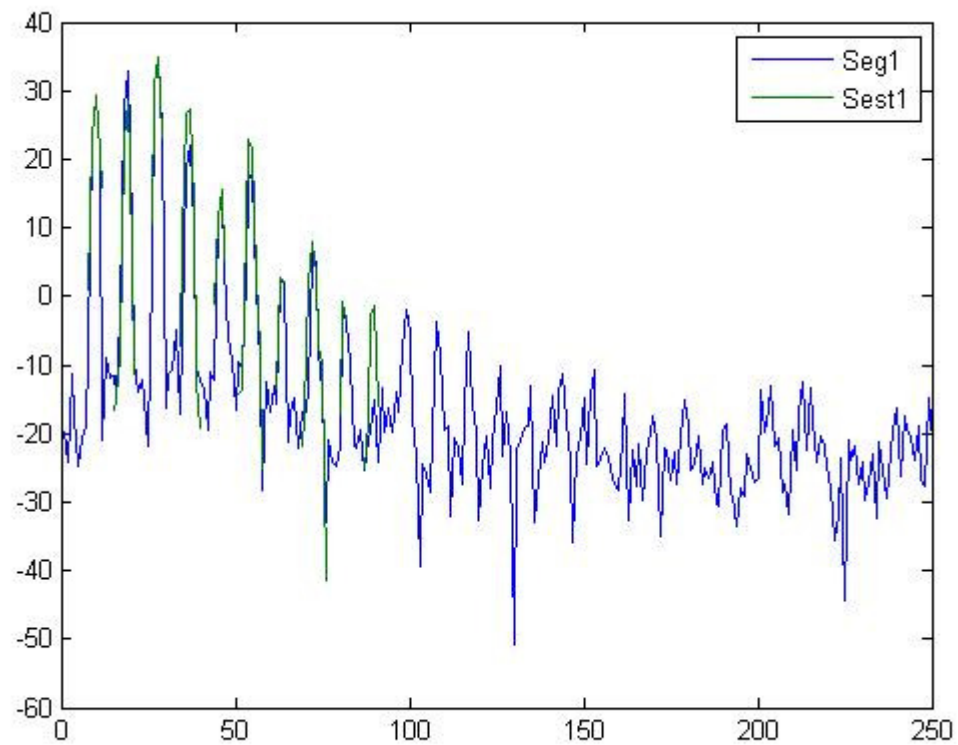
Espectro señal de trompeta nota C4:



Señal de flauta nota G4:



Espectro señal de flauta nota G4:



En el apartado 6 analizaremos los resultados y valoraremos las distintas estimaciones realizando un estudio comparativo.

6 Estudio comparativo

6.1 Introducción

En este apartado se realizará una comparación de los resultados obtenidos con los métodos estudiados en el apartado anterior:

- Modelado polinomial
- Modelado polinomial con técnicas de interpolación
- Filtro de Kalman
- Filtrado en frecuencia

La señal de test que utilizaremos tendrá las mismas características que las señales que venimos tratando en los apartados anteriores. Se han utilizado señales aisladas de distintos instrumentos musicales obtenidas de una grabación en directo, y se ha realizado la mezcla de dichas señales normalizadas, de manera artificial. De esta forma tenemos la posibilidad de medir el grado de precisión de las señales estimadas respecto a las originales. Al igual que en el apartado anterior, las características de las señales utilizadas son:

- 44100 Hz de frecuencia de muestreo
- Codificación PCM de 8 bits por muestra
- 1 canal (señal mono)
- Ruido convolutivo del propio instrumento
- Ruido aditivo (de fondo) debido al ruido ambiente de la grabación

Además de estas características técnicas, supondremos que las señales utilizadas llevan asociadas modulaciones de amplitud y frecuencia (AM+FM). Haciendo un estudio para 10 armónicos, podemos representar la señal de test de la siguiente manera:

$$s_i(n) = \sum_{i=1}^{10} A_i(n) \cos[2\pi i F_0 n + \varphi_i(n)] + r(n),$$

$$\varphi_i(n) = i A_{FM} \sin(2\pi F_{FM} n + \alpha) + \gamma_i,$$

$$A_i(n) = \frac{A_0}{i} [1 + A_{AM} \cos(2\pi F_{AM} n + \delta)].$$

Como hemos mencionado en apartados anteriores, muchas situaciones se modelan de esta forma, como por ejemplo la presencia de ‘vibrato’ en la señal, que puede definirse como una AM inducida por la FM. Gracias a esto podremos representar las variaciones en amplitud, frecuencia y fase que sufren las señales, ya que a pesar de tratarse de señales cuasi-estacionarias, es inevitable la presencia de pequeñas variaciones temporales.

A continuación se mostrará el comportamiento de los algoritmos mencionados, concretamente la estimación de señales de flauta y de trompeta. El procedimiento del estudio comparativo constará de dos partes:

- **1. Impacto del ruido:**

En la primera parte, estudiaremos el caso menos favorable: la separación dos señales separadas por un intervalo de quinta justa, y concretamente el comportamiento de los distintos métodos frente al impacto del ruido.

- **2. Calidad de la estimación en función de la desviación de frecuencias fundamentales.**

En la segunda parte, estudiaremos el comportamiento de los distintos métodos en función de la distancia entre las frecuencias fundamentales de las señales que componen la señal de test, ante las mismas condiciones generales.

El indicador que utilizaremos para medir la calidad de los algoritmos será la relación señal a residuo (SRR):

$$SRR (dB) = 10 \log_{10} \left(\frac{\sum_n s^2(n)}{E_R} \right)$$

Donde $E_R = \sum_n |s_1(n) - \widehat{s_1(n)}|^2 + \sum_n |s_2(n) - \widehat{s_2(n)}|^2$

Este es un indicador utilizado por muchos autores para medir la calidad de las estimaciones, ya que el denominador E_R recoge la energía del residuo, o lo que es lo mismo: el error de estimación de cada señal s_1 y s_2 .

6.2 Impacto del ruido

En este apartado se realizará una comparación del comportamiento de los distintos métodos estudiados, en función del impacto del ruido. Como señal de test utilizaremos la misma señal $s(n) = s1(n) + s2(n)$ que hemos tratado en el apartado 5, compuesta por:

- S1(n): Señal de Trompeta nota C4 ($F0 = 523,251$ Hz)
- S2(n): Señal de Flauta nota G4 ($F0 = 783,991$ Hz)

Las razones por las que han sido escogidas estas señales para realizar la separación de fuentes sonoras son las explicadas en el apartado 5.1, ya que al formar un intervalo de quinta justa (relación de frecuencias 3:2), nos encontramos ante el peor caso de los posibles intervalos, exceptuando el unísono (1:1) y la octava (2:1).

El procedimiento para llevar a cabo el estudio comparativo en función del impacto del ruido será añadir a la señal de test ruido blanco gaussiano de manera controlada de tal manera que obtengamos una relación señal a ruido (SNR) determinada:

$$s_{ruidosa}(n) = s_{limpia}(n) + r(n)$$

A pesar de que la señal de test $s_{limpia}(n)$ tiene incorporado ruido convolutivo del propio instrumento y ruido aditivo (de fondo) debido al ruido ambiente de la grabación, no tenemos en cuenta estos ruidos a la hora de calcular la SNR:

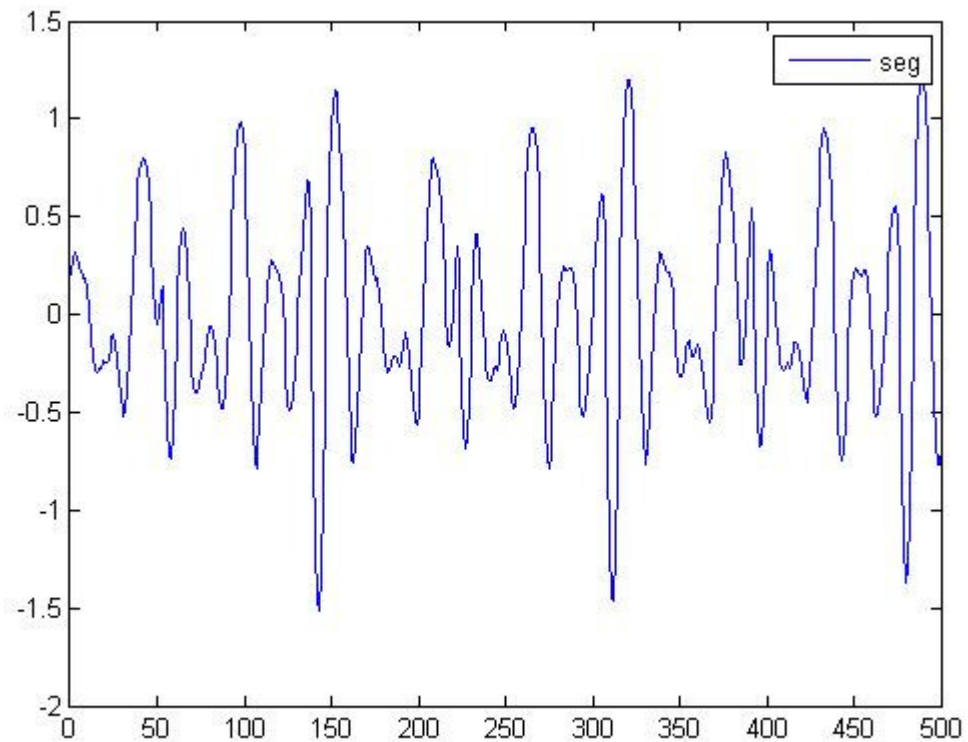
$$SNR (dB) = 10 \log_{10} \left(\frac{\sum_n s_{limpia}^2(n)}{\sum_n r^2(n)} \right)$$

De esta manera, iremos añadiendo ruido progresivamente, obteniendo valores de SNR comprendidos entre 0 y 30 dB, para comprobar el comportamiento de los distintos métodos de separación de señales, y finalmente obtener una gráfica donde se muestre la relación SRR de cada método en función de la SNR.

En las imágenes siguientes observamos el resultado de añadir ruido a la señal $s_{limpia}(n)$ mediante el procedimiento que se ha explicado, y la señal obtenida $s_{ruidosa}(n)$ con una relación señal a ruido de 0 dB:

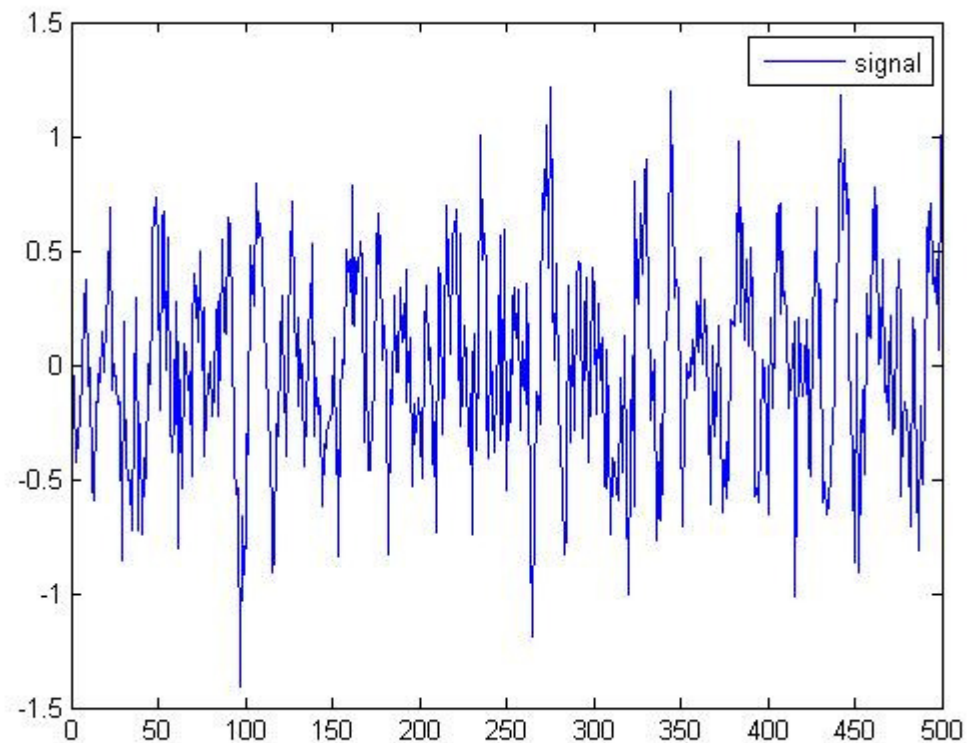
Vemos primero la señal de test original:

$S_{limpia}(n)$



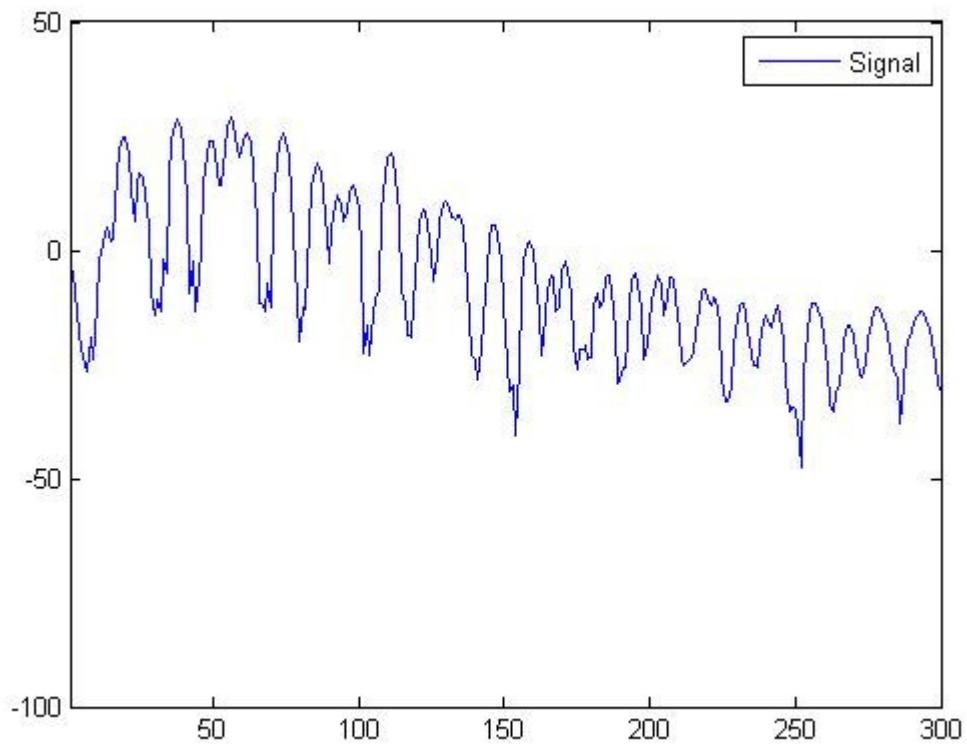
Vemos en la siguiente imagen la señal de test a la que se ha añadido ruido con una SNR = 0 dB:

$S_{ruidosa}(n)$:

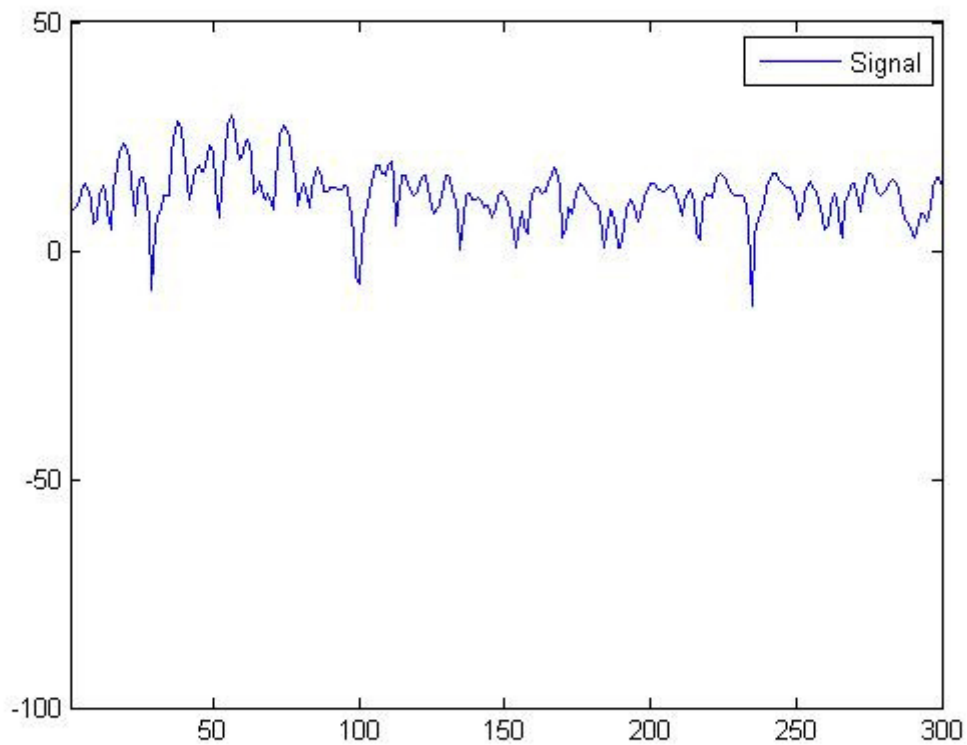


Y a continuación se muestran sus correspondientes espectros:

$S_{limpia}(n)$



$S_{ruidosa}(n)$

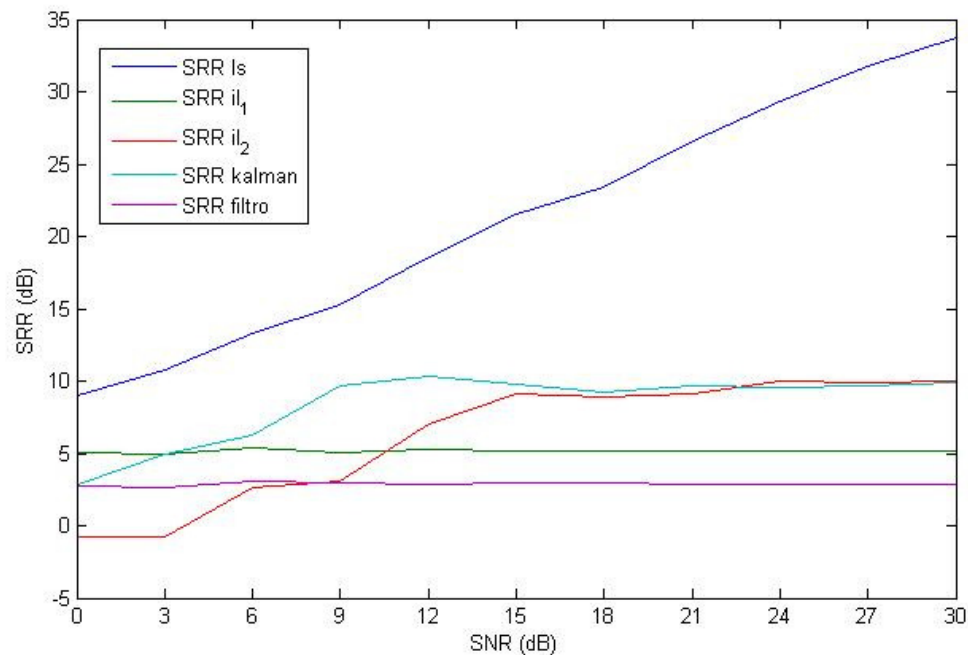


De esta manera se ha analizado el comportamiento de los distintos métodos para la señal de test con valores de SNR comprendidos entre 0 y 30 dB con saltos de 3 dB (0, 3, 6, ...30.), calculando para cada método la relación SRR en función de la SNR. Se han aplicado los algoritmos con las mismas condiciones generales, analizando tramas de 500 muestras con solapamiento del 50%. Se ha calculado el valor de SRR para una trama por cada método. En un análisis más riguroso se habría calculado el valor de SRR promedio de varias tramas, pero no ha sido posible realizarlo porque el método de interpolación en el modelado polinomial sólo se ha desarrollado para aplicarlo en tramas aisladas. Aunque de todas formas, al tratarse de señales cuasi-estacionarias, se supone que la señal se comporta de manera similar en todas sus tramas. Se ha hecho un análisis con 10 armónicos, y también se han utilizado las siguientes condiciones particulares:

- Para el modelado polinomial se ha hecho un análisis con 10 armónicos, utilizando polinomios de orden $M=4$.
- Para el filtro de Kalman se han utilizado los siguientes valores iniciales de ruido de medida, ruido de proceso y coeficientes del modelo:
 - $SIGMA = 1$ como parámetro inicial de la matriz de covarianza Q , que define el ruido de proceso.
 - $\sigma = 0,5$ como parámetro inicial del ruido de medida.
 - coeficientes = 0 como valor inicial.

En la figura siguiente se muestra, como hemos comentado, los valores de SRR en función de SNR obtenidos mediante los siguientes métodos de separación de señales:

- LS: Modelado polinomial con coeficientes obtenidos por mínimos cuadrados
- IL_2 : Modelado polinomial con coeficientes obtenidos por interpolación lineal, con base de tiempos del modelo $t=(-N/2+0.5: N/2-0.5)$
- IL_1 : Modelado polinomial con coeficientes obtenidos por interpolación lineal, con base de tiempos del modelo $t=(0: N-1)$
- Filtro de Kalman
- Filtrado en frecuencia

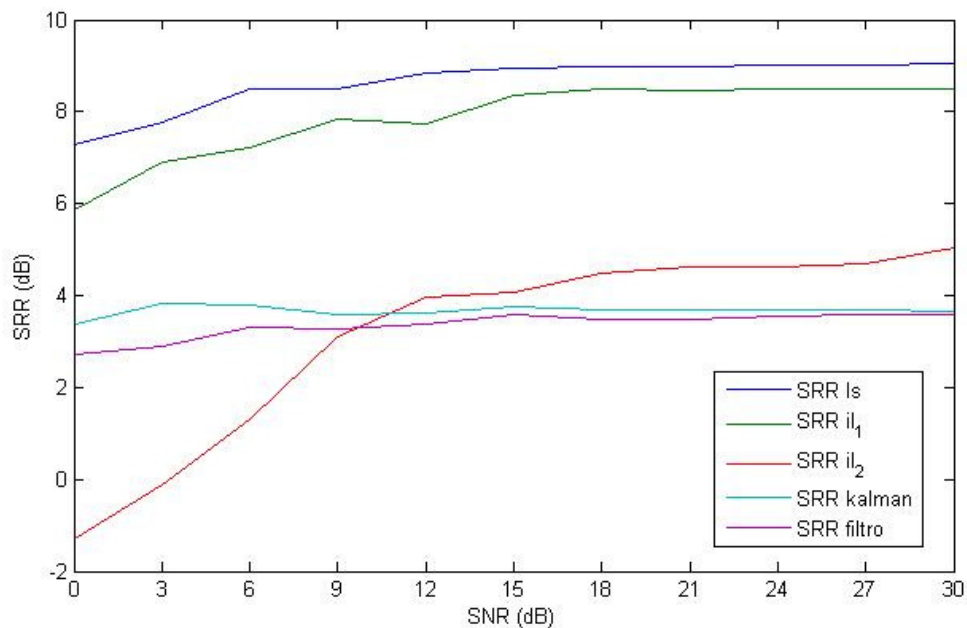


Se observa, como es lógico, que la mayor parte de los métodos estudiados obtienen mejores valores de SRR para mayores valores de SNR, a excepción del filtrado en frecuencia y el método IL1, que permanecen prácticamente invariables al impacto del ruido. A primera vista, claramente parece que el mejor método en estas condiciones es modelado polinomial con coeficientes obtenidos por mínimos cuadrados, llegando a obtener valores de SRR por encima de los 30 dB. Le siguen los métodos del filtro de Kalman y IL2, cuyos valores de SRR van aumentando conforme aumenta la SNR, hasta saturarse en torno a los 10 dB de SRR. El peor de los métodos evaluados resulta ser el filtrado en frecuencia, que permanece con $SRR = 2$ dB aproximadamente para cualquier valor de SNR.

Hay que tener en cuenta también, que las primeras muestras de las estimaciones obtenidas mediante el método del filtro de Kalman tienen un error bastante grande, debido al periodo de enganche, y hasta que se estabiliza el filtro de Kalman, no produce la estimación correcta. Se ha realizado el cálculo de SRR sin tener en cuenta esta característica, pero en el caso de no haber contabilizado estas primeras muestras del periodo de enganche, se habrían obtenido mayores valores de SRR.

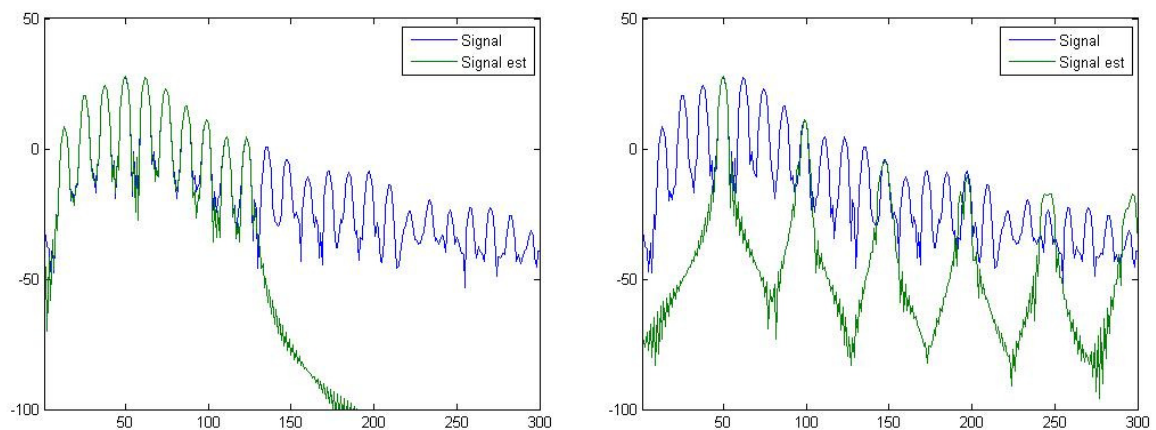
Estos resultados han sido obtenidos utilizando para todos los métodos el algoritmo de detección de frecuencia fundamental basado en el método 'ASA', empleado en apartados anteriores, y que está indicado para el caso de señales cuasi-estacionarias.

Como ya adelantamos anteriormente en el apartado 5.2, detectar correctamente la frecuencia fundamental es un paso muy importante para realizar una buena estimación. En la imagen siguiente se observan los resultados de repetir el mismo análisis utilizando el algoritmo de detección de frecuencia fundamental desarrollado por T. Grandke "Interpolation algorithms for discrete Fourier transform of weighted signals", indicado para el caso de señales con frecuencia fundamental variable en el tiempo.



Se observa que los resultados obtenidos son mucho peores que en el caso anterior. Esto es debido a la ya mencionada importancia de la detección de la frecuencia fundamental. En las figuras siguientes se muestra la diferencia entre los dos algoritmos que hemos visto:

Espectro de la señal de trompeta nota C4 (azul) y su correspondiente estimación (verde) utilizando el modelado polinomial con el método de detección de frecuencia fundamental ‘ASA’ (izquierda), y utilizando el método ‘Grandke’ (derecha):



Queda claro que en estas circunstancias el método ‘Grandke’ es mucho menos preciso, pero hay que tener en cuenta que el método ‘ASA’ suele fallar o bien en ausencia de energía de la señal (pausas entre notas) o bien si el segmento analizado es altamente no estacionario. En cambio, el método ‘Grandke’ no tiene problemas de convergencia. Por lo tanto, cada método puede tener ventajas dependiendo de la situación en la que nos encontremos.

También hay que tener en cuenta que se han realizado los cálculos utilizando un solapamiento entre tramas del 50%, si se hubiera utilizado un solapamiento mayor, los resultados de SRR habrían sido mejores, sobre todo en los métodos de interpolación.

Lo que parece quedar claro, es que el peor de los métodos evaluados en estas condiciones, resulta ser el que utiliza el filtrado en frecuencia, desarrollado por Mark R. Every y John E. Szymanski. A pesar de ser un método publicado por IEEE, consigue unos resultados bastante pobres de SRR. Esto es debido a que en su algoritmo no se tienen en cuenta las propiedades del modelado sinusoidal.

6.3 Calidad de la estimación en función de la desviación de frecuencias fundamentales

En la segunda parte del estudio comparativo, analizaremos el comportamiento de los distintos métodos de separación de fuentes sonoras en función de la distancia entre las frecuencias fundamentales de las señales que componen la señal de test, ante las mismas condiciones generales.

Para realizar el estudio, se ha calculado SRR en función de la desviación frecuencial entre las 2 señales que componen la señal de test $s(n) = s1(n) + s2(n)$, donde:

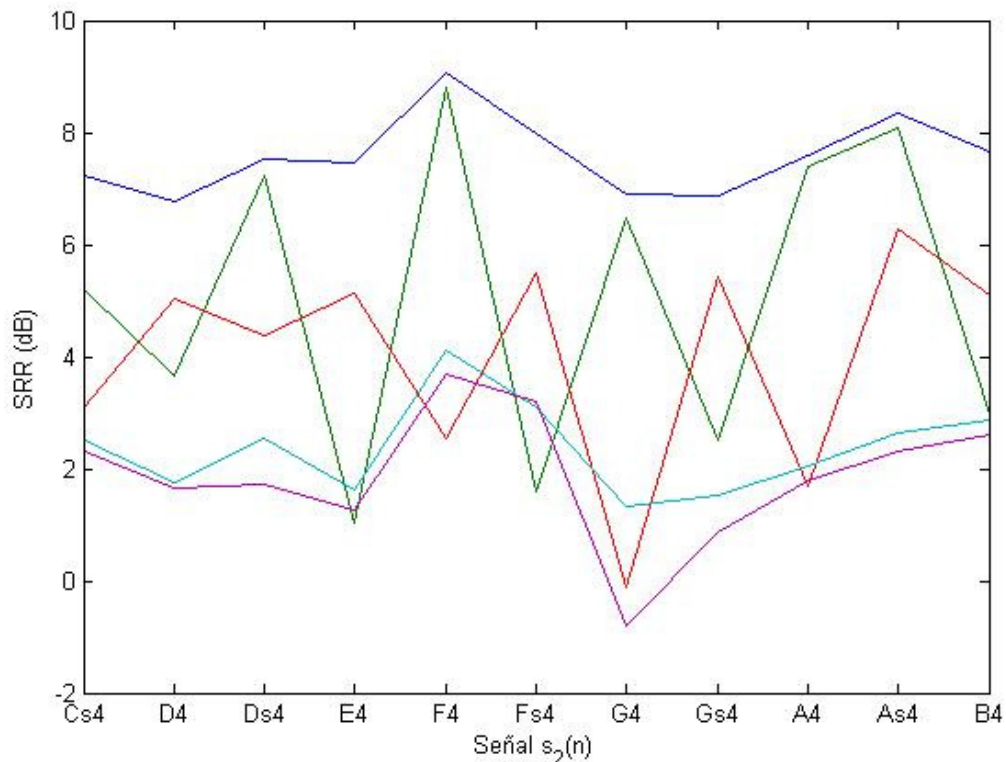
- S1(n): Señal de trompeta nota C4 ($F_0 = 523,251$ Hz)
- S2(n): Distinta señal de flauta con notas que van variando desde C#4 hasta B4 con frecuencias fundamentales:
 - C# 4: 554,365 Hz
 - D 4: 587,33 Hz
 - D# 4: 622,254 Hz
 - E 4: 659,255 Hz
 - F 4: 698,456 Hz
 - F# 4: 739,989 Hz
 - G 4: 783,991 Hz
 - G#4: 830,609 Hz
 - A 4: 880 Hz
 - A# 4: 932,328 Hz
 - B 4: 987,767 Hz

Se ha calculado el valor de SRR para una trama por cada método en condiciones de $SNR = 30$ dB, y se han aplicado los algoritmos con las mismas condiciones generales, analizando tramas de 500 muestras con solapamiento del 50%, y analizando un solo armónico. También se han utilizado las siguientes condiciones particulares:

- Para el modelado polinomial se ha hecho un análisis con un solo armónico (el fundamental), utilizando polinomios de orden $M=4$.
- Para el filtro de Kalman se han utilizado los siguientes valores iniciales de ruido de medida y ruido de proceso:
 - SIGMA = 1 como parámetro inicial de la matriz de covarianza Q, que define el ruido de proceso.
 - sigma = 0,5 como parámetro inicial del ruido de medida.
 - coeficientes = 0 como valor inicial.

En la figura siguiente se muestra, como hemos comentado, los valores de SRR en función de la desviación de las frecuencias fundamentales de las señales que componen la señal de test, obtenidos mediante los siguientes métodos de separación de señales:

- LS: Modelado polinomial con coeficientes obtenidos por mínimos cuadrados
- IL_2 : Modelado polinomial con coeficientes obtenidos por interpolación lineal, con base de tiempos del modelo $t=(-N/2+0.5: N/2-0.5)$
- IL_1 : Modelado polinomial con coeficientes obtenidos por interpolación lineal, con base de tiempos del modelo $t=(0: N-1)$
- Filtro de Kalman
- Filtrado en frecuencia



Estos resultados han sido obtenidos utilizando para todos los métodos el algoritmo de detección de frecuencia fundamental basado en el método ‘ASA’, empleado en apartados anteriores, y que está indicado para el caso en el que nos encontramos de señales cuasi-estacionarias.

Se observa en general, que el peor caso para la mayor parte de los métodos es el de la quinta justa (C4+G4), como era de esperar por lo explicado en el apartado 5.1. También se

aprecia que la mayor parte de los métodos siguen la misma tendencia tomando valores alrededor de 8 dB (LS) y 2 dB (Filtro de Kalman y Filtrado en frecuencia), a excepción de los métodos que obtienen los coeficientes por interpolación (IL_1 e IL_2), para los cuales es difícil deducir alguna dependencia con la desviación de frecuencias fundamentales.

De la misma manera que en el apartado anterior, aquí también hay que tener en cuenta que las primeras muestras de las estimaciones obtenidas mediante el método del filtro de Kalman tienen un error bastante grande, debido al periodo de enganche, y hasta que se estabiliza el filtro de Kalman, no se produce la estimación correcta. Al realizar el estudio sólo con el armónico fundamental, este efecto parece afectar más a la estimación. Al igual que en el apartado anterior, se ha realizado el cálculo de SRR sin tener en cuenta esta característica, pero en el caso de no haber contabilizado estas primeras muestras del periodo de enganche, se habrían obtenido mayores valores de SRR.

De todas formas, queda claro que el peor método en estas circunstancias, sigue siendo el filtrado en frecuencia, desarrollado por Mark R. Every y John E. Szymanski.

7 Conclusiones y líneas futuras

7.1 Conclusiones

Al margen de los objetivos concretos del presente proyecto, cualquier estudio o proyecto de investigación tiene como objetivo principal llegar a ciertas conclusiones sobre el tema estudiado. Recordando los objetivos que nos planteábamos al abordar este proyecto en el apartado 3, buscábamos realizar la separación de señales de forma correcta, especialmente para armónicos fuertemente solapados en frecuencia, puesto que los métodos existentes no proporcionan una solución satisfactoria en estas circunstancias. De esta manera, a lo largo del proyecto hemos trabajado en la separación de señales cuasi-armónicas mediante el estudio y comparación de varios métodos alternativos.

En la primera parte del proyecto hemos trabajado con señales cuasi-estacionarias procedentes de una sola fuente, y en la segunda parte hemos analizado señales más complejas, compuestas por señales procedentes de varias fuentes y con armónicos fuertemente solapados en frecuencia, comprobando el funcionamiento de los distintos métodos de separación de señales.

Dentro del estudio del modelado polinomial, hemos llevado a cabo un estudio de los coeficientes del modelo, realizando una estimación de dichos coeficientes mediante técnicas de interpolación. Esta técnica, basada en el criterio de continuidad temporal de los coeficientes para una fuente sonora cuasi-estacionaria, ha sido desarrollada con intención de resolver la problemática planteada en el apartado 2.1: realizar una correcta separación de fuentes sonoras cuasi-armónicas. En conclusión, podemos decir que hemos introducido pequeñas modificaciones en el algoritmo, dando como resultado dos métodos nuevos que hemos denominado abreviadamente IL_1 e IL_2 , que posteriormente evaluamos en el estudio comparativo.

En el estudio del comportamiento del filtro de Kalman, hemos descubierto que gran parte del error de estimación del algoritmo es debido a las primeras muestras de cada trama, que están dentro del llamado periodo de enganche del filtro de Kalman, y hasta que se estabiliza no se produce la estimación correcta. Además de esto, hemos comprobado que el buen funcionamiento de este algoritmo depende en gran medida de la correcta elección de los valores iniciales del ruido de medida y el ruido de proceso.

También hemos estudiado el método desarrollado por Mark R. Every y John E. Szymanski basado en el filtrado en frecuencia, descubriendo que es el que peores resultados obtiene de todos los comparados, a pesar de ser un método publicado por IEEE. Esto es debido a que el algoritmo no hace uso de las ventajas del modelado sinusoidal.

Además de los problemas concretos que pueda tener cada método a la hora de realizar una correcta separación de señales cuasi-armónicas, hemos comprobado que un paso crítico para todos los algoritmos es la correcta detección de frecuencia fundamental. Se ha trabajado con dos métodos de detección de frecuencia fundamental: ‘ASA’: automated spectral analysis, y ‘Grandke’, y hemos descubierto que el primero realiza una detección de F_0 mejor para señales cuasi-estacionarias, pero suele fallar o bien en ausencia de energía de la señal (pausas entre notas) o bien si el segmento analizado es altamente no estacionario. En cambio, el método ‘Grandke’ no tiene problemas de convergencia. Por lo

tanto, cada método de detección de F0 puede tener ventajas dependiendo de la situación en la que nos encontremos.

Después de analizar lo problemática que puede llegar a ser la separación de fuentes sonoras cuasi-armónicas, y más concretamente el caso de separación de señales separadas por un intervalo de quinta justa, al tener varios armónicos solapados en frecuencia, hemos comprobado el funcionamiento de los distintos métodos de separación de señales ante las mismas condiciones generales.

Por último, hemos realizado un estudio comparativo para medir la calidad de los distintos algoritmos ante las mismas circunstancias.

La primera parte de este estudio comparativo analizaba el impacto del ruido en la calidad de la separación de fuentes sonoras cuasi-armónicas (separadas por un intervalo de quinta justa), y se ha comprobado que la mayor parte de los métodos estudiados obtienen mejores valores de SRR para mayores valores de SNR, como era de esperar. También se ha visto que el mejor método de los estudiados es el modelado polinomial y el peor el filtrado en frecuencia.

En la segunda parte del estudio comparativo hemos analizado el comportamiento de los distintos métodos de separación de fuentes sonoras en función de la distancia entre las frecuencias fundamentales de las señales que componen la señal de test, ante las mismas condiciones generales. Con este estudio hemos comprobado que el peor caso para la mayor parte de los métodos es el de la quinta justa (C4+G4), ya que las señales que componen la señal de test tienen una relación de frecuencias fundamentales 3:2, y debido a esto varios de los armónicos de las señales se solapan. Este era un resultado esperado por lo explicado en el apartado 5.1, y de hecho por esa razón hemos utilizado la quinta justa como señal de test en todo el apartado 5, para evaluar el comportamiento de los métodos de separación ante señales cuasi-armónicas, ya que al formar un intervalo de quinta justa (relación de frecuencias 3:2), nos encontramos ante el peor caso de los posibles intervalos, exceptuando el unísono (1:1) y la octava (2:1) que no entran en el estudio.

Hay que mencionar también, que la medida de calidad utilizada en el estudio comparativo ha sido la relación señal a residuo, que no tiene en cuenta la aplicación a la que va dirigida la separación. Pero también se podían haber utilizado otros indicadores que se basen en otros principios, como por ejemplo la audibilidad, u otras medidas de calidad específicas para aplicaciones concretas.

Como conclusión final, podemos decir que hemos comprobado el correcto funcionamiento de distintos métodos de separación de fuentes sonoras, pero el rendimiento de los algoritmos existentes suele ser bastante limitado, sobretudo en el caso de que se presenten armónicos fuertemente solapados en frecuencia. Además, hemos descubierto a qué se deben las distintas limitaciones generales y particulares de cada método, intentando encontrar soluciones para mejorar su funcionamiento.

7.2 Líneas futuras

En este proyecto se han mostrado algunas de las técnicas que se pueden utilizar para realizar la separación de fuentes sonoras, y se ha comprobado que el mayor problema a la hora de llevar a cabo la separación es la presencia de fuentes cuasi-armónicas. Ante estas circunstancias, se ha descubierto a qué se deben las limitaciones particulares de cada método estudiado, y se ha intentando encontrar soluciones para mejorar el funcionamiento de los mismos. En este camino, se han dejado muchas puertas abiertas por las que se podría seguir investigando para encontrar mejoras en los algoritmos. A continuación enumeraremos varias de las posibles líneas futuras de investigación enfocadas a cada método en concreto:

- *Modelado polinomial con técnicas de interpolación*

En el apartado 4 hemos estudiado el criterio de continuidad temporal de los coeficientes para una fuente sonora cuasi-estacionaria, llegando a desarrollar un método de estimación que hemos denominado abreviadamente IL_1 . El funcionamiento de este método consiste en aproximar el valor de los coeficientes del modelo polinomial de una trama concreta, a través de la interpolación lineal de los coeficientes (a_0 y b_0) de las tramas adyacentes. En este sentido se podría investigar con otros tipos de interpolación que posiblemente obtengan valores más precisos.

En ese mismo apartado también se ha visto que había otra forma de realizar la interpolación, y es a través del coeficiente $A = \sqrt{a_0^2 + b_0^2}$, que representa la amplitud de cada parcial. Con este procedimiento se obtienen valores mucho más precisos, pero tiene el inconveniente de que los coeficientes del modelo (a_0 y b_0) permanecen desconocidos, haciendo imposible la reconstrucción de la señal. Esto nos da la idea de que otra posible línea futura de trabajo podría ser no utilizar los valores obtenidos por interpolación como solución, sino como condiciones restrictivas del sistema de ecuaciones.

También en el apartado 4 hemos observado la gran variabilidad de los coeficientes de mayor orden (a_3 y b_3), que adquieren valores muy grandes perjudicando la precisión de la interpolación. En este sentido, hemos experimentado con la modificación de los exponentes y la base de tiempos del modelo, con intención de que los coeficientes tomen valores más pequeños para que así la interpolación logre obtener valores más precisos. Hemos llegado a la conclusión de que un modelo con exponentes negativos produce que los coeficientes tomen valores mucho más pequeños, pero los resultados que se obtienen al realizar la estimación de la señal con estas modificaciones no son nada satisfactorios, al producirse inestabilidades. Otra posible línea futura de trabajo podría ir encaminada a controlar dichas inestabilidades.

- *Filtro de Kalman*

Hemos comprobado el buen funcionamiento del filtro de Kalman en la estimación de señales, detectando que sus limitaciones son debidas principalmente al periodo de enganche del filtro de Kalman, que no produce una estimación correcta hasta que no se estabiliza. Una clara línea de investigación podría ir en este sentido. Como hemos

adelantado en el apartado 4.7, parece claro que este efecto podría ser mitigado haciendo que las tramas sean más largas, y esta ha sido una de las opciones estudiadas para intentar mejorar el funcionamiento del algoritmo. Pero las modificaciones que hemos realizado no han dado buenos resultados, debido al hecho de que el filtro de Kalman requiere que los coeficientes del modelo se vayan reiniciando en cada trama para hacer una buena predicción.

Además de esto, hemos comprobado que el buen funcionamiento de este algoritmo depende en gran medida de la correcta elección de los valores iniciales del ruido de medida y el ruido de proceso. Otra posible línea de trabajo podría analizar pormenorizadamente los posibles valores iniciales para encontrar la mejor solución.

8 Bibliografía

En la realización de este proyecto se utilizaron las siguientes referencias:

- [1] Miroslav Zivanovic and Johan Schoukens, “Time-variant harmonic signal modelling by using polynomial approximation and fully automated spectral analysis”, Proceedings of the 17th European Signal Processing Conference, Glasgow, UK, August 2009.
- [2] Tuomas Virtanen, “Sound Source Separation in Monaural Music Signals” Thesis for the degree of Doctor of Technology to be presented with due permission for public examination and criticism in Tietotalo Building, Auditorium TB109, at Tampere University of Technology, on the 3rd of November 2006, at 12 noon.
- [3] Xavier Xerra, “Musical Sound Modeling with Sinusoids plus Noise”, published in C. Roads, S. Pope, A. Piccialli, G. De Poli, editors. 1997. “Musical Signal Processing”. Swets & Zeitlinger Publishers.
- [4] Özgül Salor, Mübeccel Demirekler and Umut Orguner: “Kalman Filter Approach for Pitch Determination of Speech Signals”, SPECOM'2006, St. Petersburg, 25-29 June 2006
- [5] Hamid Satar-Boroujeni and Bahram Shafai, “Tracking of Partial in Music Signals Using Kalman Filtering: Modeling and Analysis”
- [6] Ali Taylan Cemgil, Bert Kappen and Peter Desain, Henkjan Honing, “On tempo tracking: Tempogram Representation and Kalman filtering”
- [7] Mark R. Every and John E. Szymanski “Separation of Synchronous Pitched Notes by Spectral Filtering of Harmonics”, IEEE Transactions on audio, speech and language processing, vol 14, n° 5, September 2006.



ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
INDUSTRIALES Y DE TELECOMUNICACIÓN



SEPARACIÓN DE FUENTES SONORAS CUASI-ARMÓNICAS

PFC Ing. Telecomunicación

Iñigo Andión Iracheta

Tutor: Miroslav Zivanovic

Pamplona, 20 de Julio de 2011

Motivación

- Las señales acústicas del mundo real normalmente están compuestas por la suma de señales procedentes de distintas fuentes sonoras.

$$x(n) = \sum_{m=1}^M s_m(n), \quad n = 1, \dots, N$$

- Existen varios métodos para realizar la separación , pero su **rendimiento es limitado**, especialmente cuando las fuentes son cuasi-armónicas.

Objetivos

- Estudio y comparación de distintos métodos existentes de separación de señales.
- Comportamiento en presencia de armónicos fuertemente solapados.
- Propuesta de modificaciones en los algoritmos para solucionar deficiencias.

Índice

- Introducción
 - Problemática
 - Aplicaciones
 - Estado del Arte
- Análisis de señales procedentes de **una fuente**:
 - Modelado polinomial
 - Resultados
 - Estudio de los coeficientes del modelo
 - Filtro de Kalman
 - Resultados
 - Estudio de los coeficientes del modelo

Índice

- Separación de señales cuasi-armónicas procedentes de **varias fuentes**:
 - Modelado polinomial
 - Modelado polinomial con técnicas de interpolación
 - Filtro de Kalman
 - Filtrado en frecuencia
- Estudio comparativo:
 - Impacto del ruido
 - Calidad en función de la desviación de F_0
- **Conclusiones** y líneas futuras

Introducción: Problemática

- Existen varios métodos para realizar la separación, pero su **rendimiento es limitado**.
- Los algoritmos fallan cuando las fuentes sonoras son **cuasi-armónicas**.
- La calidad de la estimación depende en gran medida de la **información previa conocida**.
- Estudiaremos el comportamiento de varios métodos en el escenario del **peor caso**:
 - Señales monofónicas
 - Armónicos solapados
 - Sin utilizar información previa conocida

Introducción: Aplicaciones

- En **procesado de audio**:
 - Si tenemos una señal polifónica, resulta imposible aplicar cualquier tratamiento individualmente a alguna de las fuentes que la componen y no al conjunto.
 - Si se realiza la separación de señales correctamente, se abren **multitud de aplicaciones**.

$$s(n) = s_1(n) + s_2(n)$$

Introducción: Aplicaciones

- En **procesado de audio**:
 - Transcripción automática de partituras
 - Procesado independiente de las fuentes que componen una señal polifónica:
 - Remezcla de canciones
 - Karaoke
- Aplicaciones en **otros campos**:
 - Mecánica
 - Medicina

Introducción: Estado del arte

1. Métodos basados en el **modelado sinusoidal**:

- Señal = Parciales (Determinista) + Ruido (Residuo)

$$x(n) = \sum_{h=1}^H a_h \cos(2\pi f_h n / f_s + \theta_h) + r(n), \quad n = 0, \dots, N-1,$$

2. Aprendizaje no supervisado:

- Señal = suma ponderada de funciones base b_j :

$$\hat{x}_t = \sum_{j=1}^J g_{j,t} b_j \quad t = 1, \dots, T,$$

Introducción: Estado del arte

3. Métodos basados en características psicoacústicas:

– Asociar componentes con sus correspondientes fuentes sonoras basándose en el funcionamiento del **cerebro humano**:

- Proximidad espectral
- Concordancia armónica
- Cambios síncronos de los componentes
- Proximidad espacial

Análisis señales simples

- Análisis de señales procedentes de **una fuente**:
 - Modelado polinomial
 - Resultados
 - Estudio de los coeficientes del modelo
 - Filtro de Kalman
 - Resultados
 - Estudio de los coeficientes del modelo

Modelado polinomial

- Hace uso del **modelado sinusoidal**:

$$s(n) = \sum_{i=1}^I A_i(n) \cos[2\pi i F_0(n)n + \varphi_i] + r(n), \quad n = 0, 1, \dots, N-1.$$

- Aplicando identidades trigonométricas:

$$s(n) = \sum_{i=1}^I a_i(n) \sin(2\pi i F_0(n)n) + b_i(n) \cos(2\pi i F_0(n)n) + r(n),$$

$$\hat{a}_i(n) = \sum_{m=0}^M a_m^{(i)} n^m, \quad \hat{b}_i(n) = \sum_{m=0}^M b_m^{(i)} n^m,$$

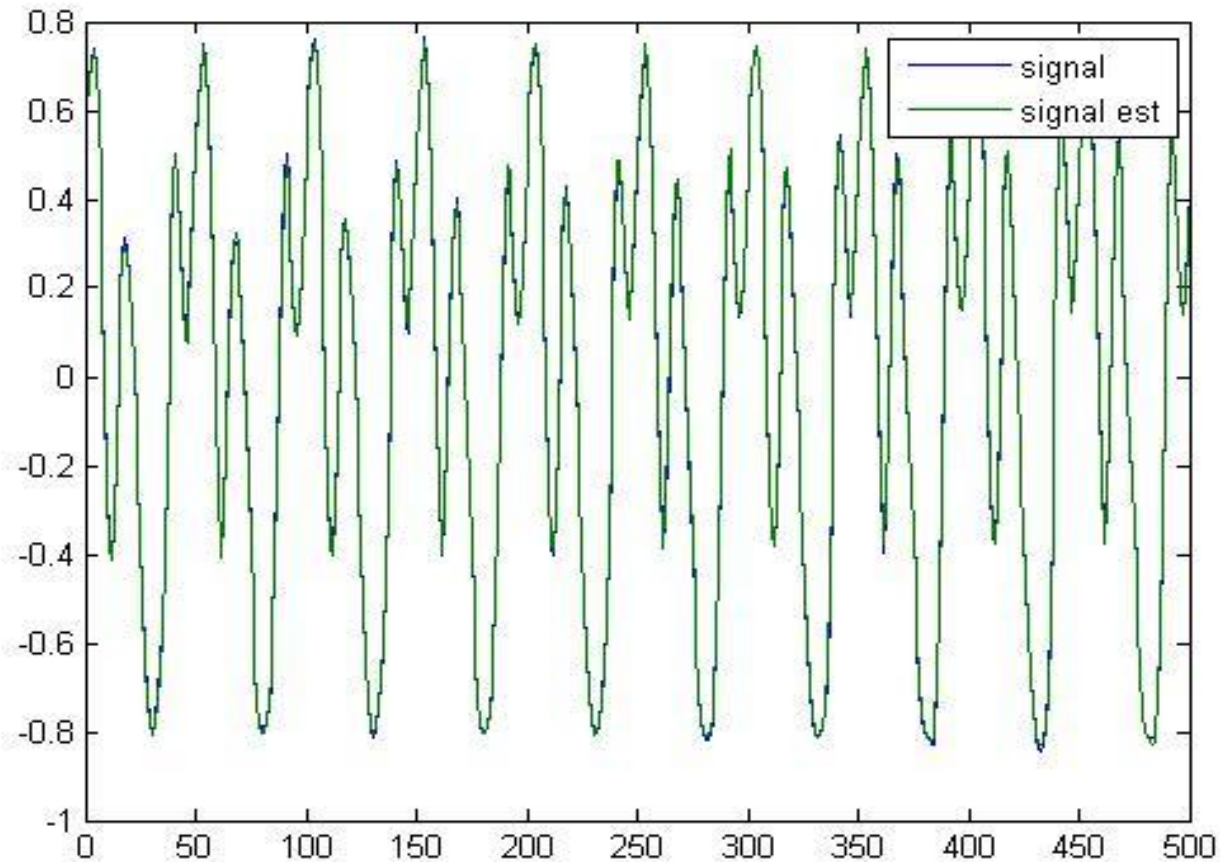
Modelado polinomial

- Primer paso: **detección de F_0**
 - Método 'ASA': $F_0 = 1/(\text{distancia entre picos})$
- Obtención del resto de los parámetros del modelo mediante mínimos cuadrados (**LS**).

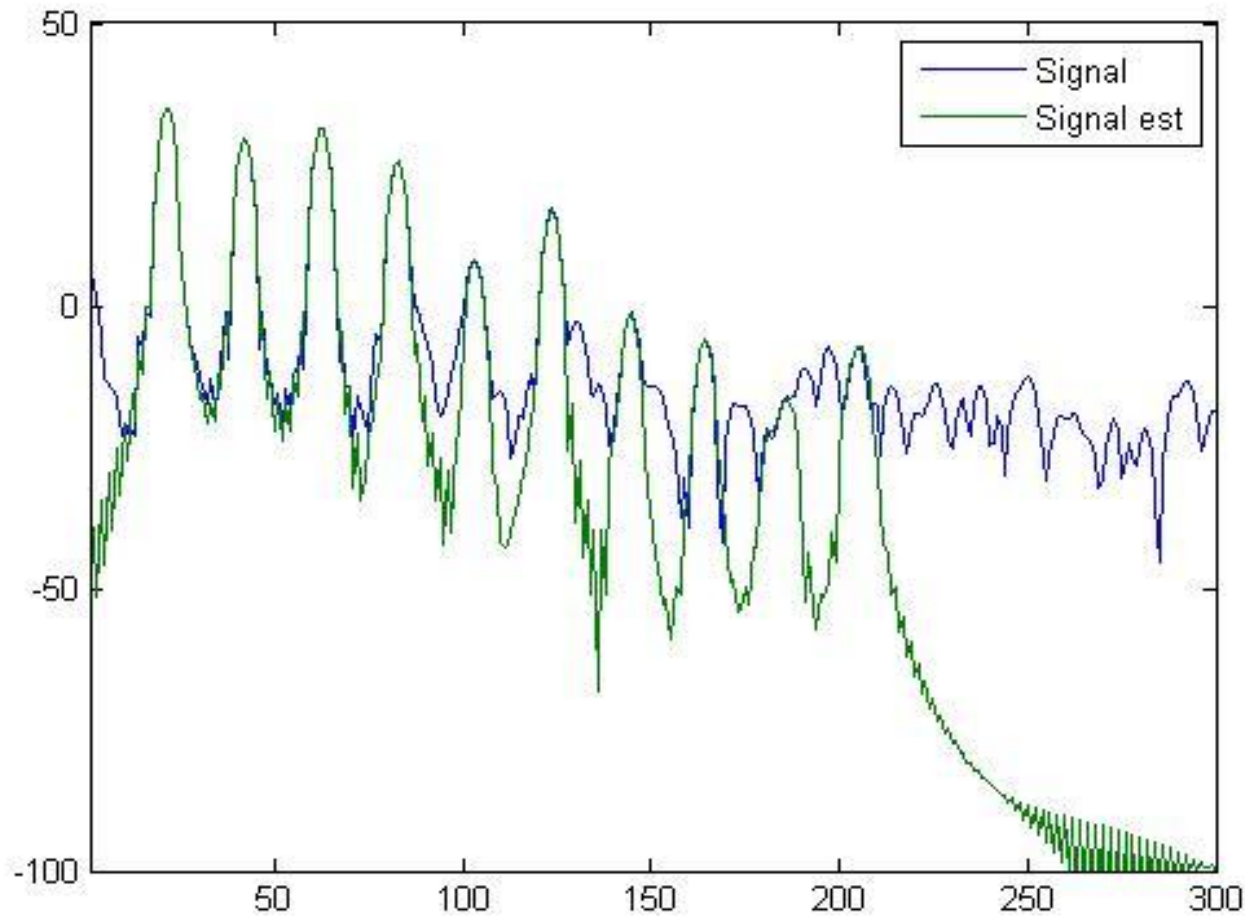
$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\sin 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\cos 2\pi f_0n$$

- Estudio realizado para 10 armónicos, con polinomios de orden 4

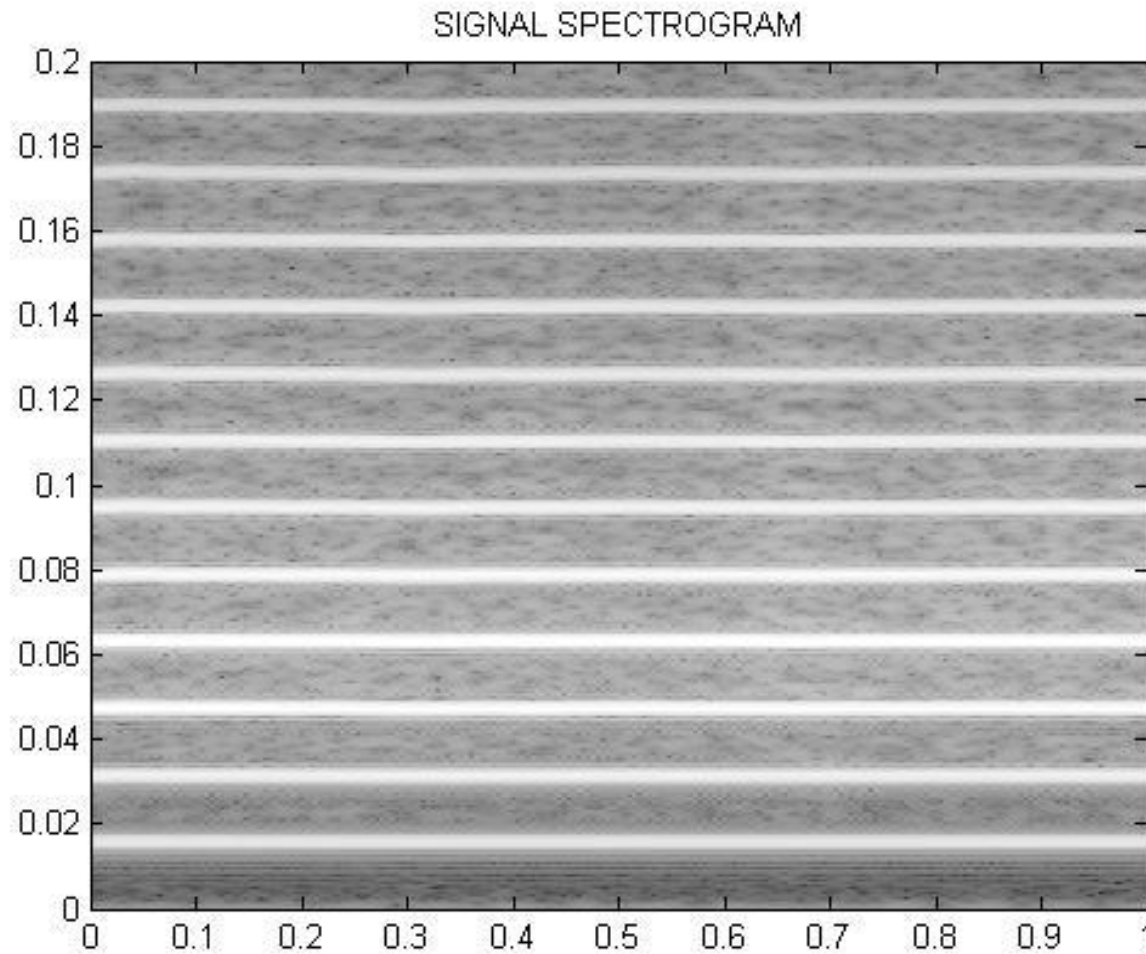
Modelado polinomial



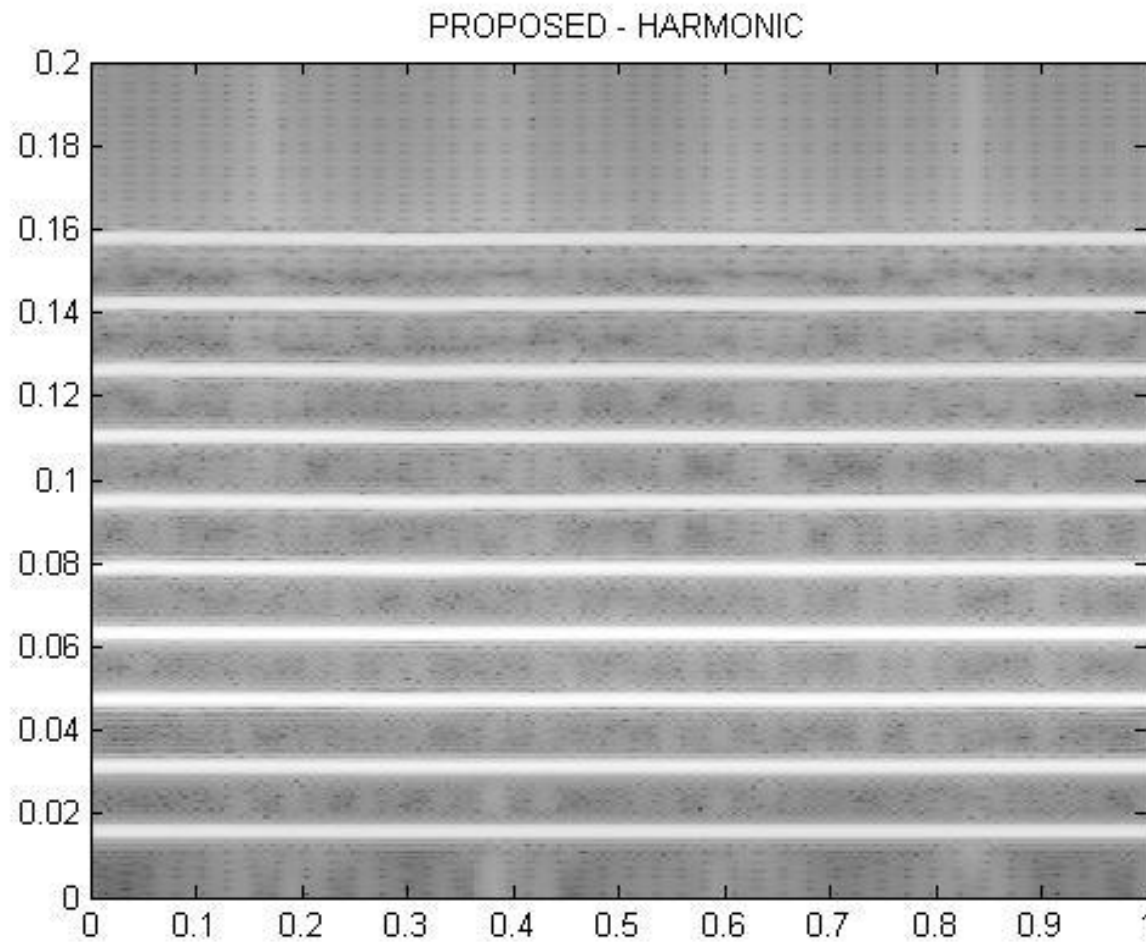
Modelado polinomial



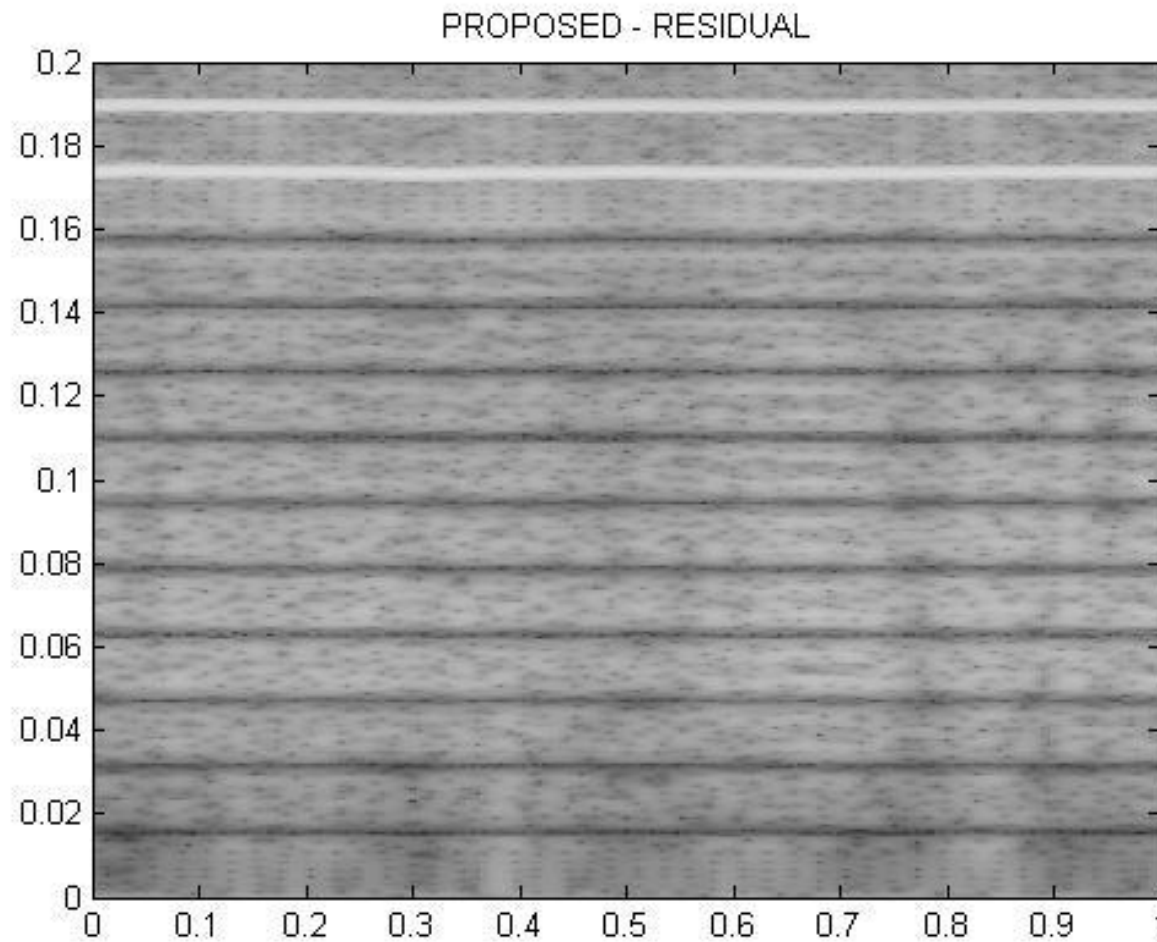
Modelado polinomial



Modelado polinomial

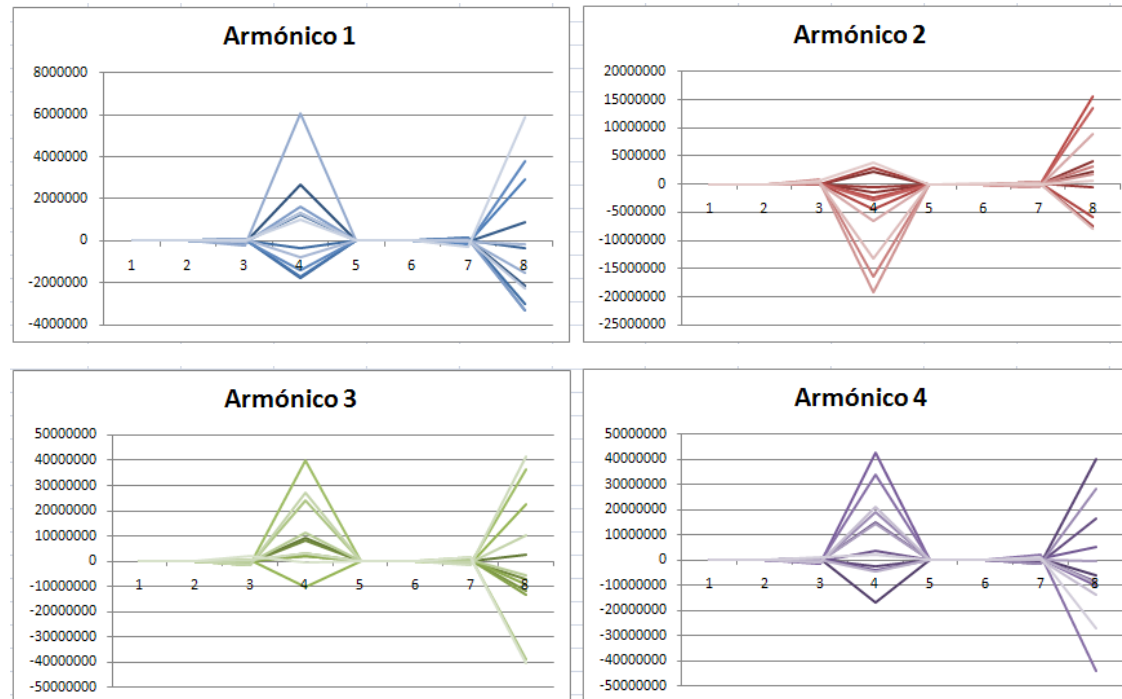


Modelado polinomial



Modelado polinomial

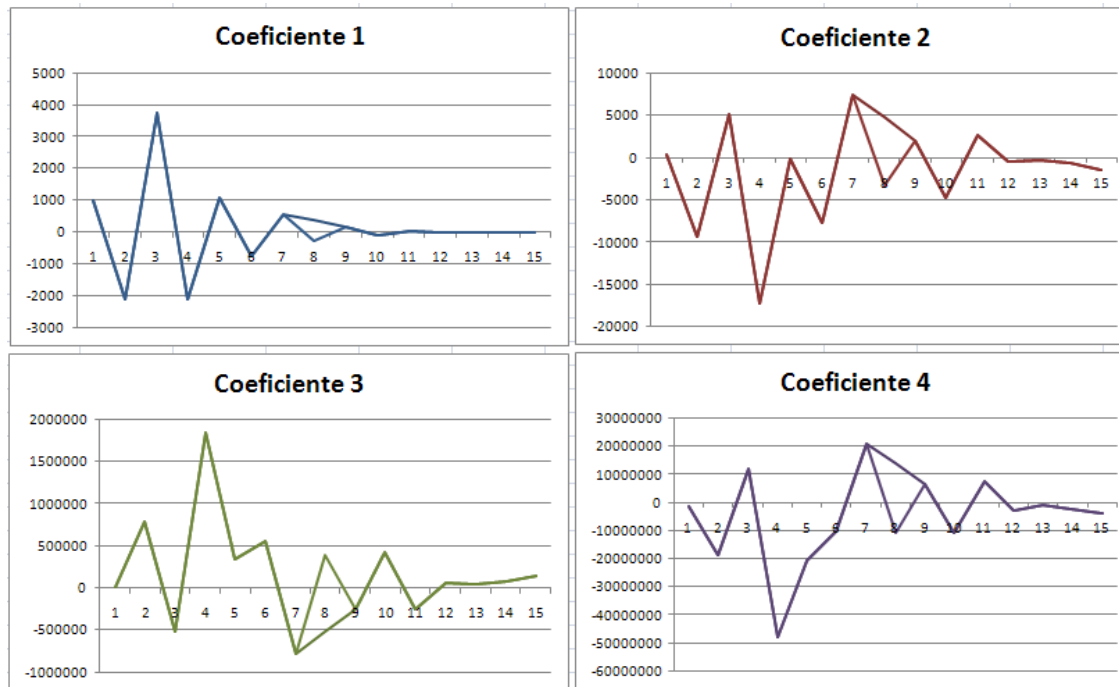
- Análisis de los coeficientes de un solo armónico en muchas tramas:



$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\text{sen } 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\text{cos } 2\pi f_0n$$

Modelado polinomial

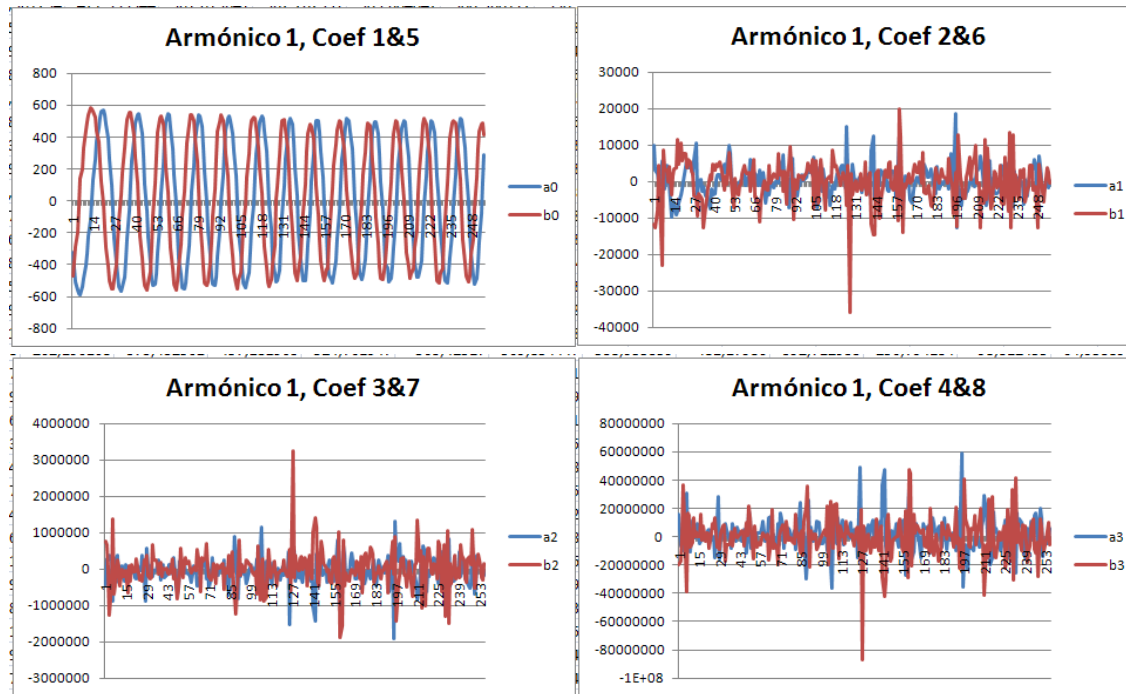
- Análisis del mismo coeficiente para todos los armónicos:



Deducción de los coeficientes del armónico 8 a partir de los armónicos vecinos mediante interpolación lineal.

Modelado polinomial

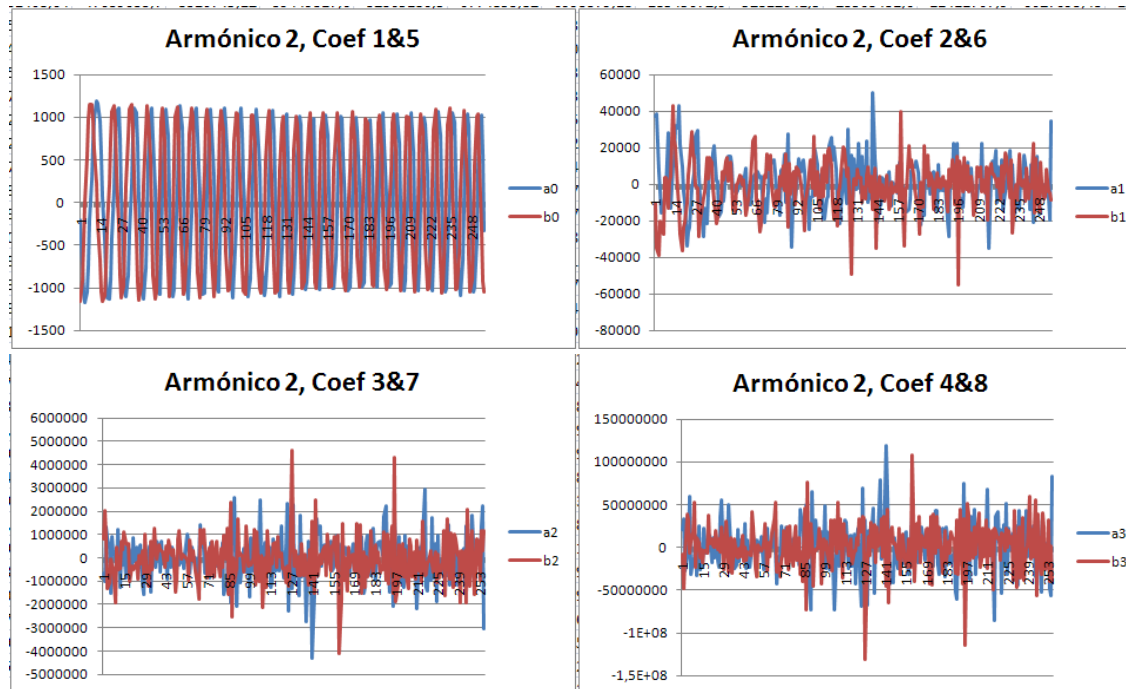
- Evolución temporal de los coeficientes:



$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\sin 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\cos 2\pi f_0n$$

Modelado polinomial

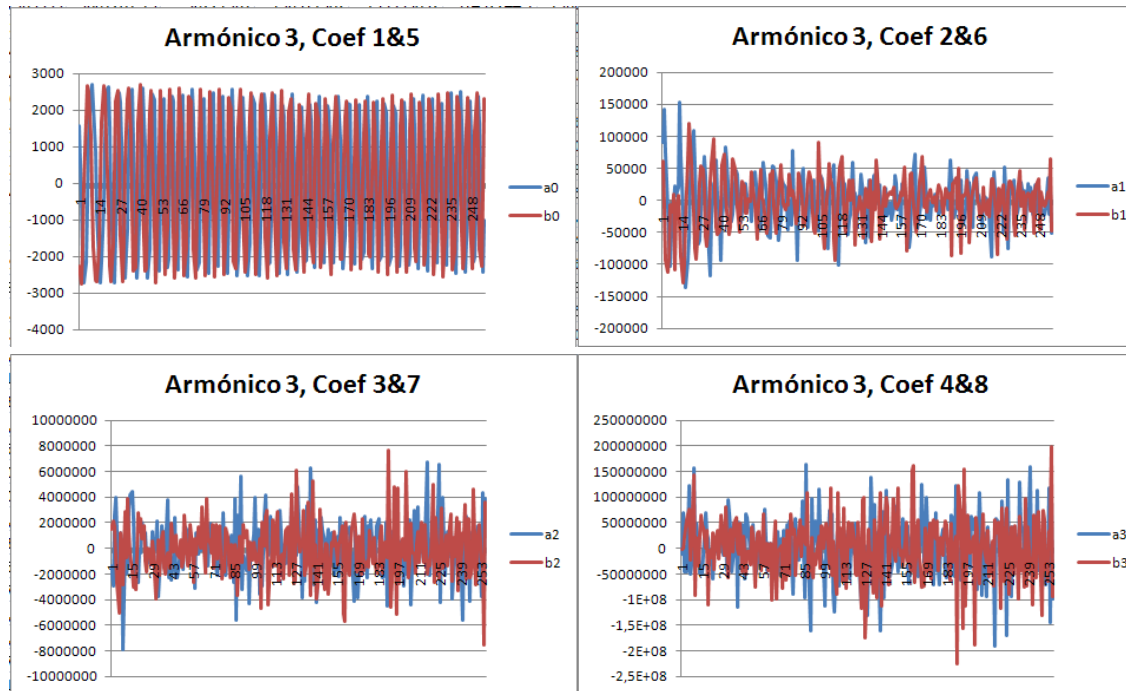
- Evolución temporal de los coeficientes:



$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\sin 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\cos 2\pi f_0n$$

Modelado polinomial

- Evolución temporal de los coeficientes:

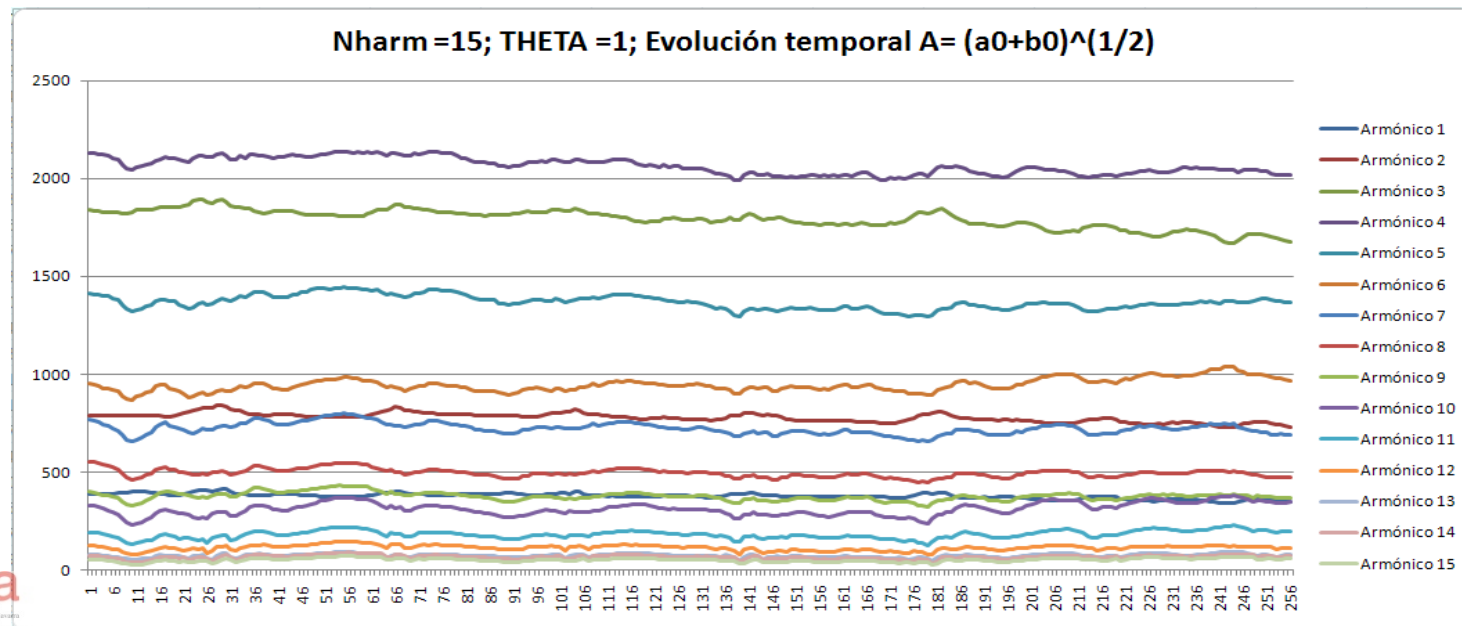


$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\sin 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\cos 2\pi f_0n$$

Modelado polinomial

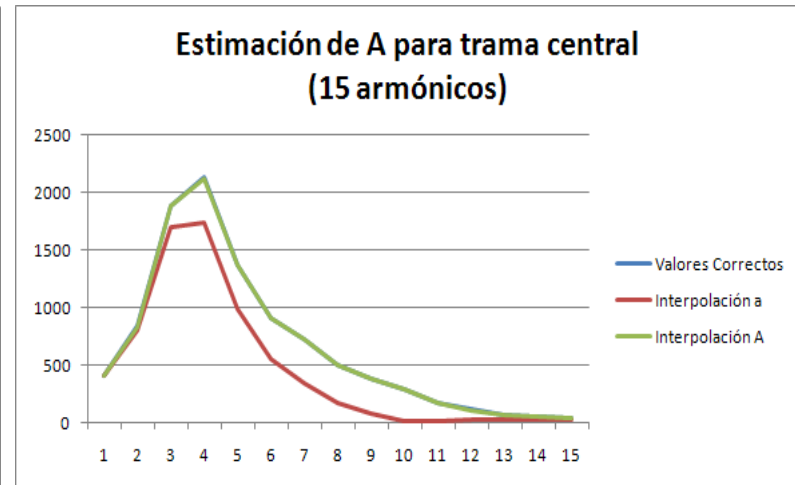
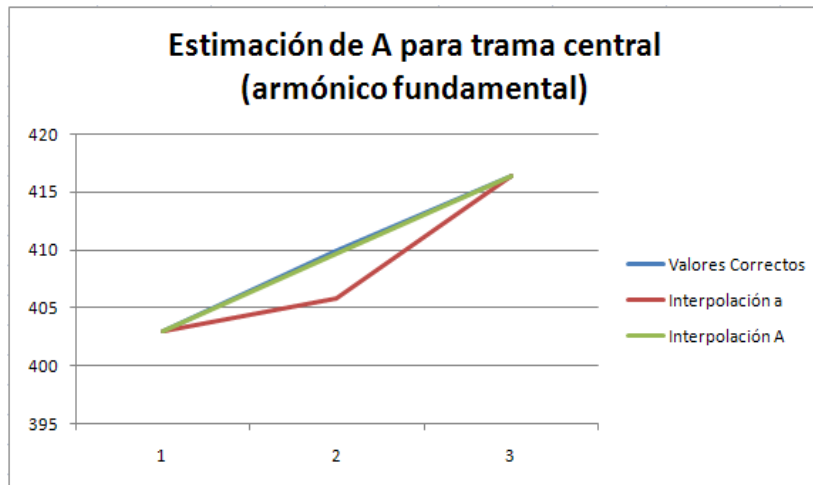
- Evolución temporal de los coeficientes (señal cuasi-estacionaria):
 - los coeficientes (a_0, b_0) siguen un patrón periódico
 - el resto de coeficientes toman valores impredecibles
- La amplitud de cada armónico: $A = \sqrt{a_0^2 + b_0^2}$

$$s(n) = (a_0) \sin 2\pi f_0 n + (b_0) \cos 2\pi f_0 n$$



Modelado polinomial

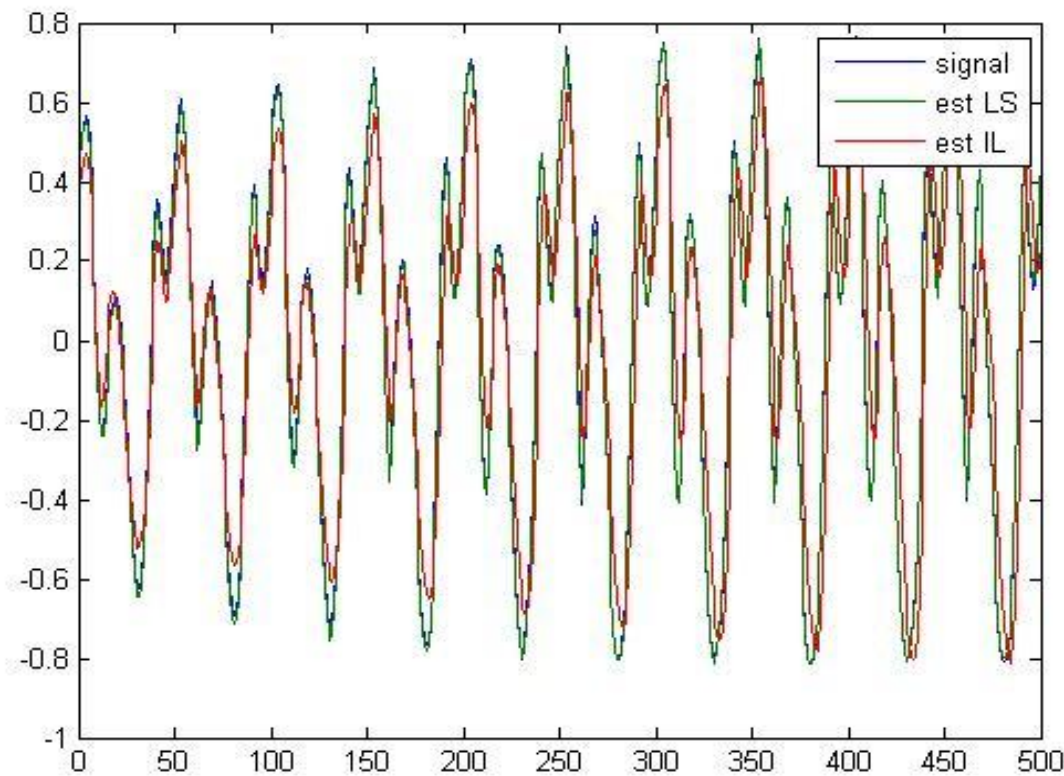
- Estimación de los coeficientes del modelo para una trama a partir de los coeficientes de las tramas adyacentes.



- rojo, se obtiene por interpolación $(a_0 \text{ y } b_0)$, y a partir de estos se calcula $A = \sqrt{a_0^2 + b_0^2}$
- verde, se obtiene directamente el valor de A de la trama central, por interpolación de los valores de A de las tramas vecinas

Modelado polinomial

- Reconstrucción de la señal utilizando los valores de la interpolación:



Modelado polinomial

- Modificación de la base de tiempos del modelo
 - Modelo exponentes positivos:

$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\text{sen } 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\text{cos } 2\pi f_0n$$

- Modelo exponentes negativos:

$$s(n) = (a_0 + a_1n^{-1} + a_2n^{-2} + a_3n^{-3})\text{sen } 2\pi f_0n + (b_0 + b_1n^{-1} + b_2n^{-2} + b_3n^{-3})\text{cos } 2\pi f_0n$$

- Bases de tiempos (valores de n):

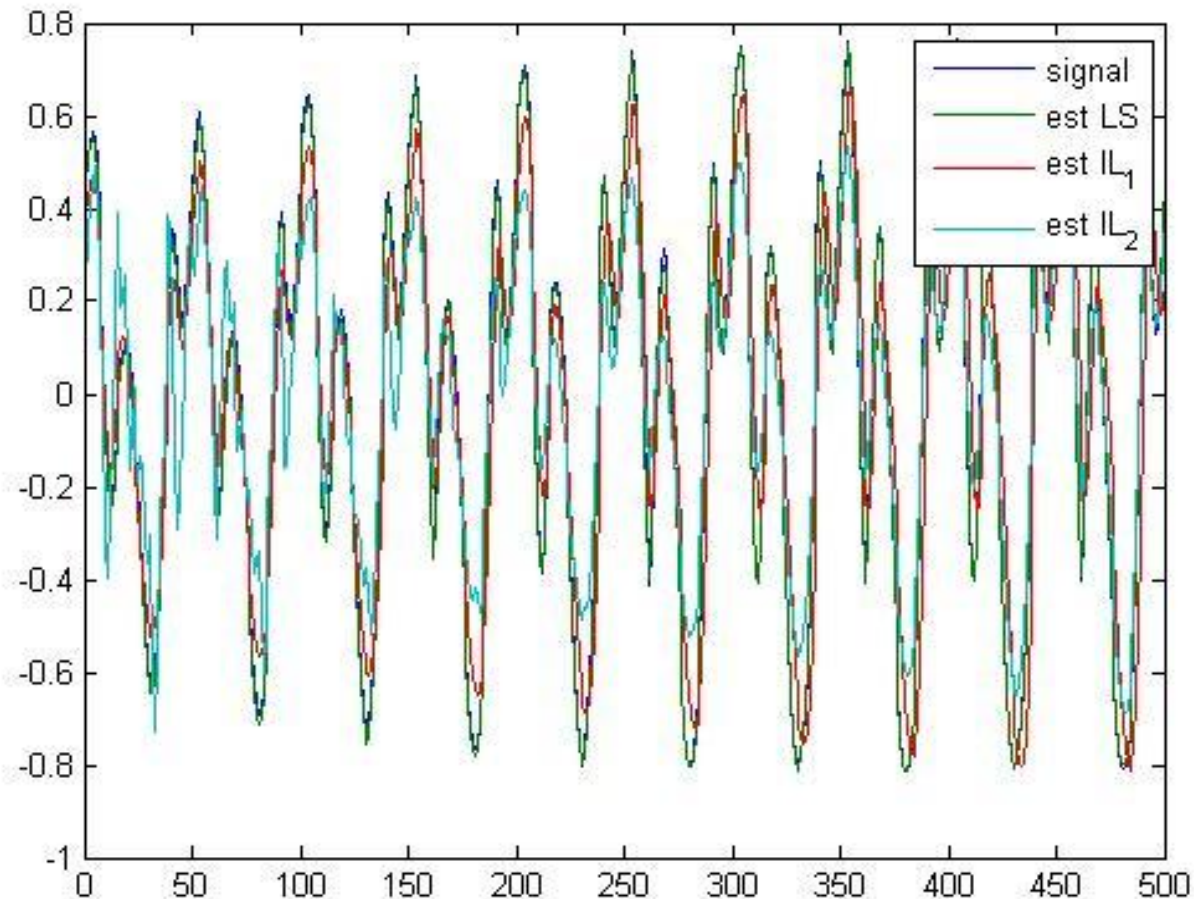
$$t = (0 : N-1)$$

$$t = (-N/2+0.5 : N/2-0.5)$$

$$t = (N : 2N-1)$$

$$t = (2N : 3N-1)$$

Modelado polinomial



Filtro de Kalman

- Hace uso del **modelado sinusoidal**:

$$s(n) = \sum_{i=1}^I A_i(n) \cos[2\pi i F_0(n)n + \varphi_i] + r(n), \quad n = 0, 1, \dots, N-1.$$

- Y también utiliza la aproximación polinomial:

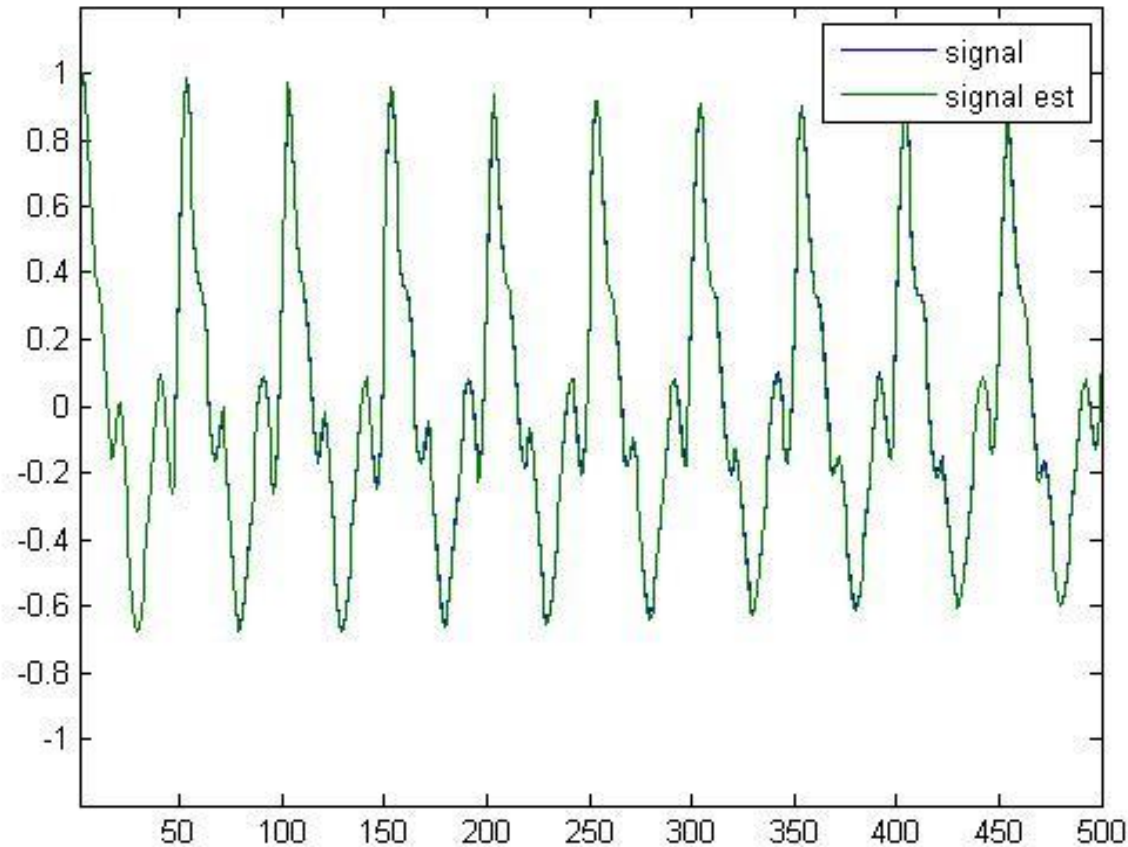
$$s(n) = \sum_{i=1}^I a_i(n) \sin(2\pi i F_0(n)n) + b_i(n) \cos(2\pi i F_0(n)n) + r(n),$$

$$\hat{a}_i(n) = \sum_{m=0}^M a_m^{(i)} n^m, \quad \hat{b}_i(n) = \sum_{m=0}^M b_m^{(i)} n^m,$$

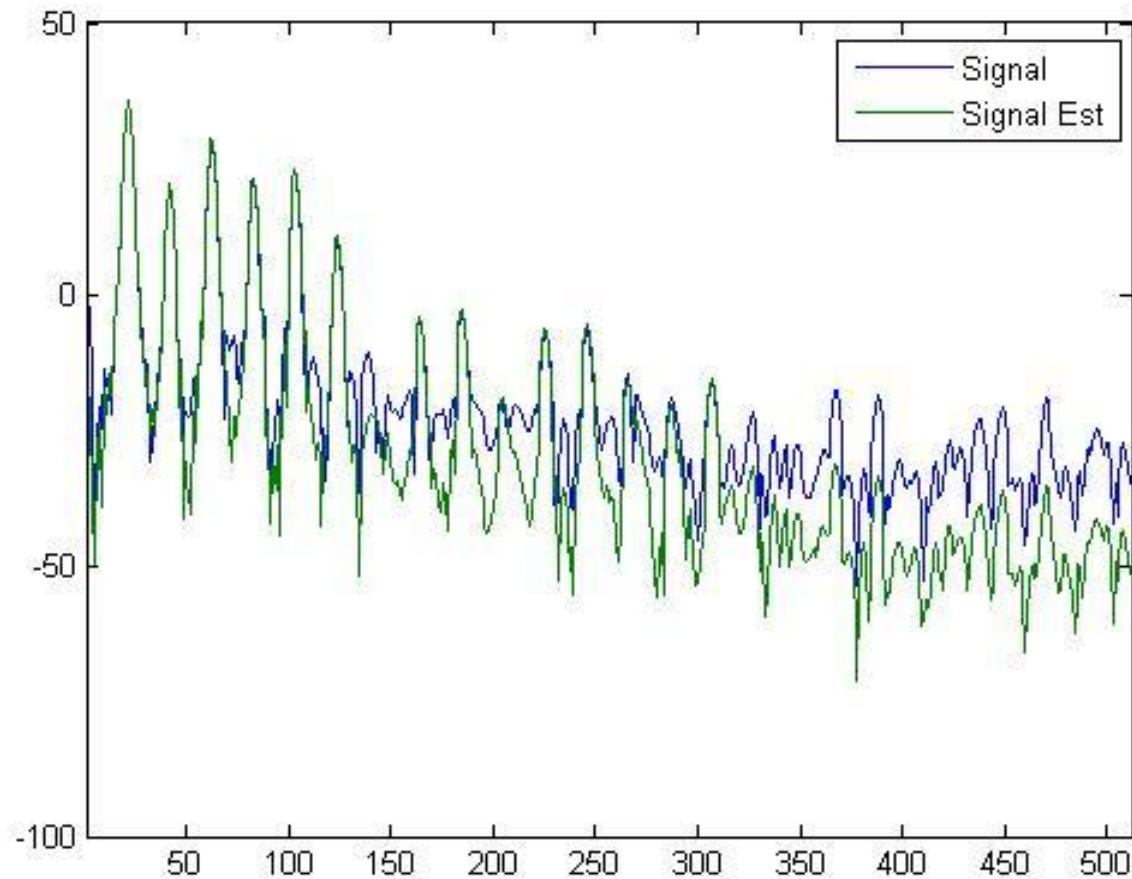
Filtro de Kalman

- Obtiene los coeficientes del modelo mediante el algoritmo del **filtro de Kalman**, que descompone la parte residual en dos tipos de ruido:
 - Ruido de medida: Debido al error de observación.
 - Ruido proceso: Debido al error del modelo.
- Necesario inicializar estos valores, además de los coeficientes del modelo.

Filtro de Kalman

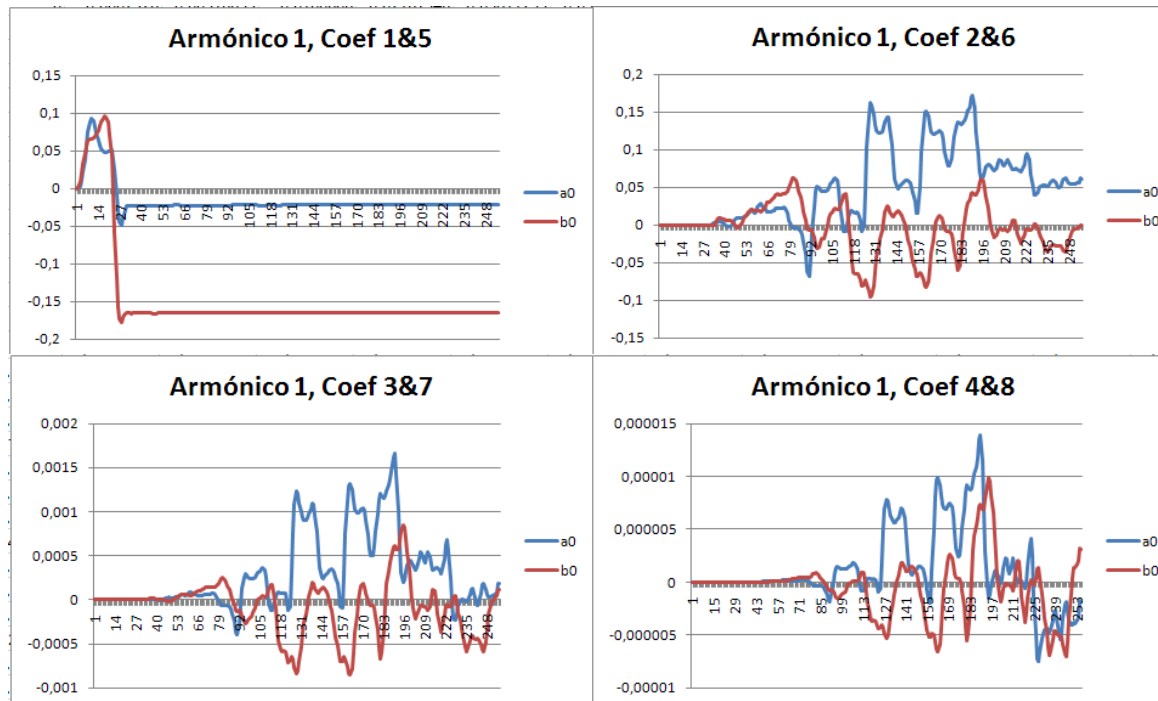


Filtro de Kalman



Filtro de Kalman

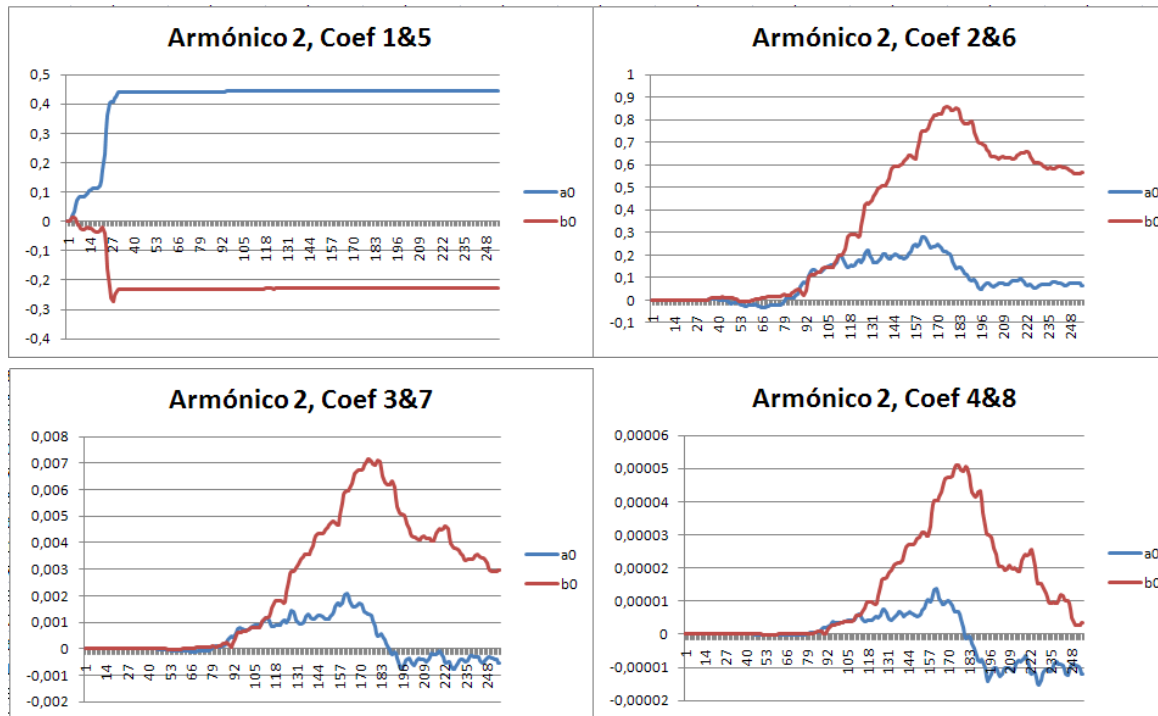
- Evolución temporal de los coeficientes (1 trama):



$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\sin 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\cos 2\pi f_0n$$

Filtro de Kalman

- Evolución temporal de los coeficientes (1 trama):



$$s(n) = (a_0 + a_1n + a_2n^2 + a_3n^3)\text{sen } 2\pi f_0n + (b_0 + b_1n + b_2n^2 + b_3n^3)\text{cos } 2\pi f_0n$$

Separación de señales cuasi-armónicas

$$s(n) = s_1(n) + s_2(n)$$

- Estudiaremos el comportamiento de:
 - Modelado polinomial
 - Modelado polinomial con técnicas de interpolación
 - Filtro de Kalman
 - Filtrado en frecuencia

Separación de señales cuasi-armónicas

$$s(n) = s_1(n) + s_2(n)$$

- S1 = Señal de Trompeta nota C4 (F0 = 523,251 Hz)
- S2 = Señal de Flauta nota G4 (F0 = 783,991 Hz)

Intervalo de **quinta justa** (relación de frecuencias 3:2)

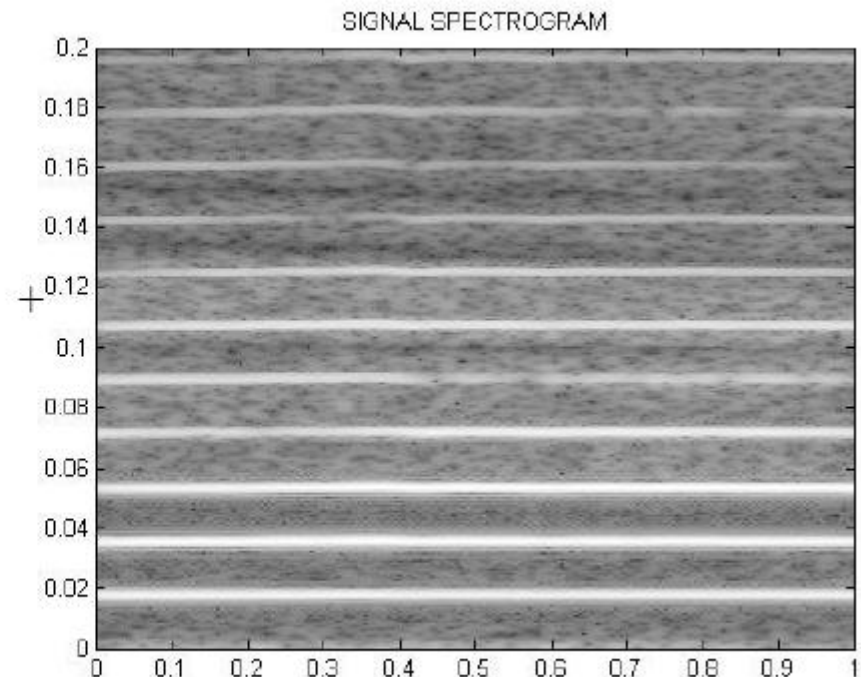
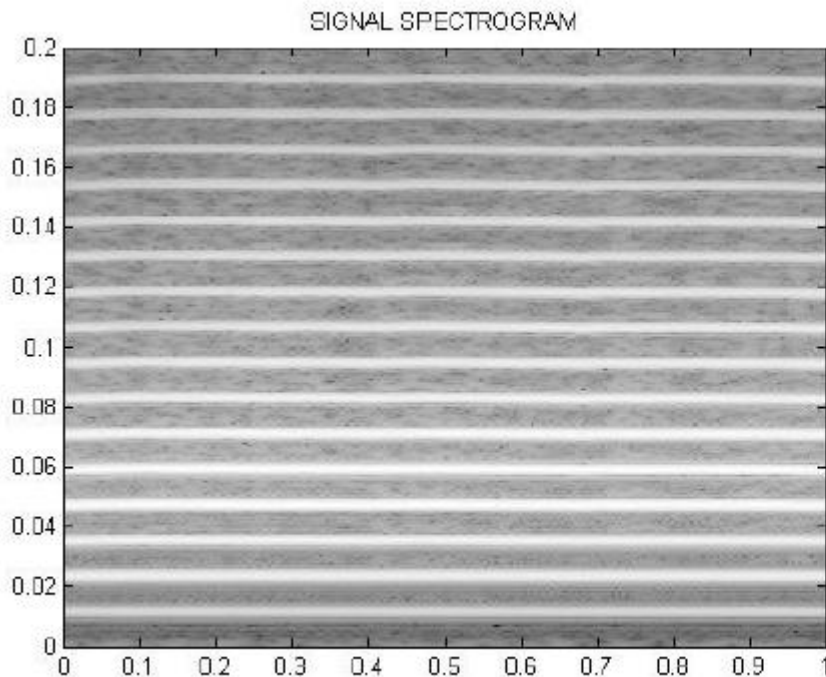
 Comparten gran nº de armónicos

Separación de señales

				HARMÓNICOS							
Intervalo	Relación	1	2	3	4	5	6	7	8	9	10
Unísono	1:1	100.0	200.0	300.0	400.0	500.0	600.0	700.0	800.0	900.0	1000.0
Octava Justa	2:1	200.0	400.0	600.0	800.0	1000.0					
Quinta Justa	3:2	150.0	300.0	450.0	600.0	750.0	900.0				
Quarta Justa	4:3	133.3	266.6	400.0	533.3	666.6	800.0	933.3			
Sexta Mayor	5:3	166.6	333.2	500.0	666.6	833.3	1000.0				
Tercera Mayor	5:4	125.0	250.0	375.0	500.0	625.0	750.0	875.0	900.0		
Tercera menor	6:5	120.0	240.0	360.0	480.0	600.0	720.0	840.0	960.0		
Segunda Mayor	9:8	112.5	225.0	337.5	450.0	562.5	675.0	787.5	900.0		
Sexta menor	8:5	166.0	332.0	498.0	664.0	830.0	996.0				

Separación de señales

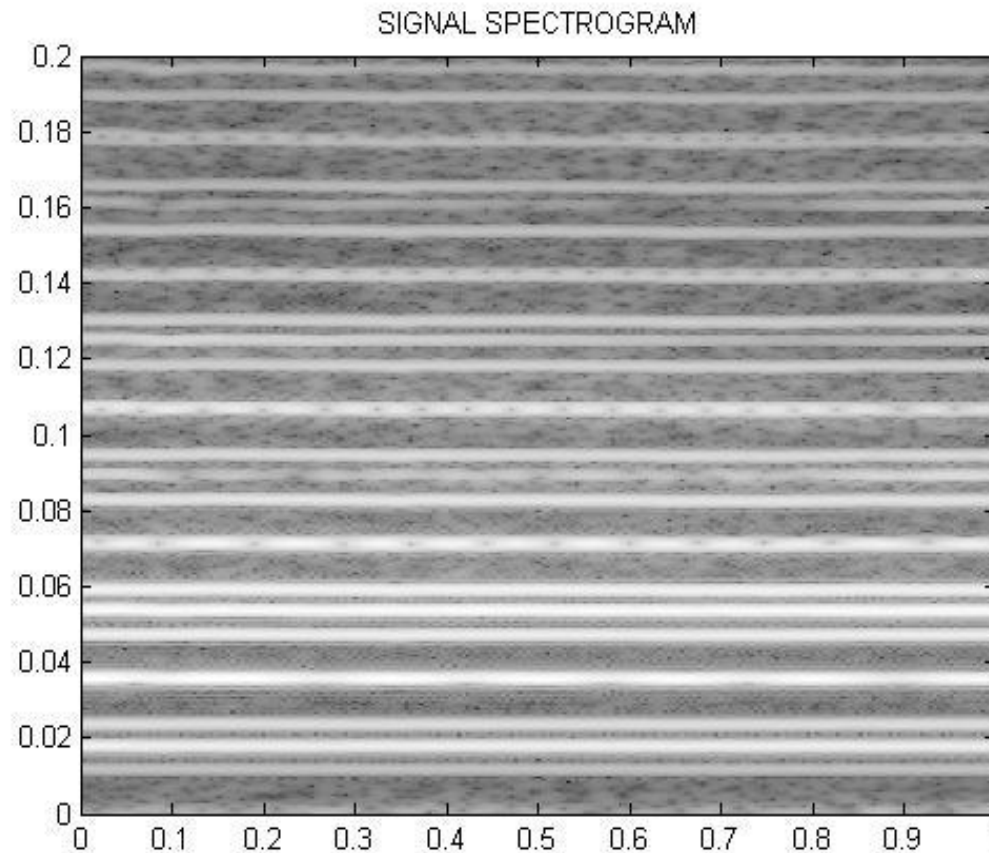
- Señal de test:



Armónicos 2º, 4º, y 6º de la señal de la derecha se solapan con los armónicos 3º, 6º y 9º de la señal de la izquierda

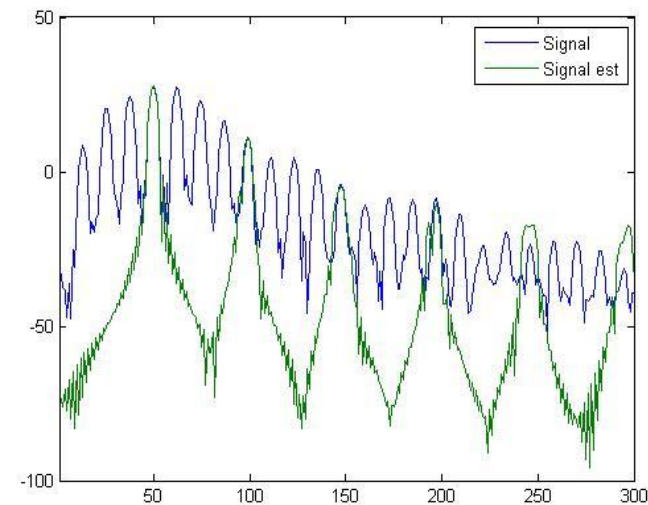
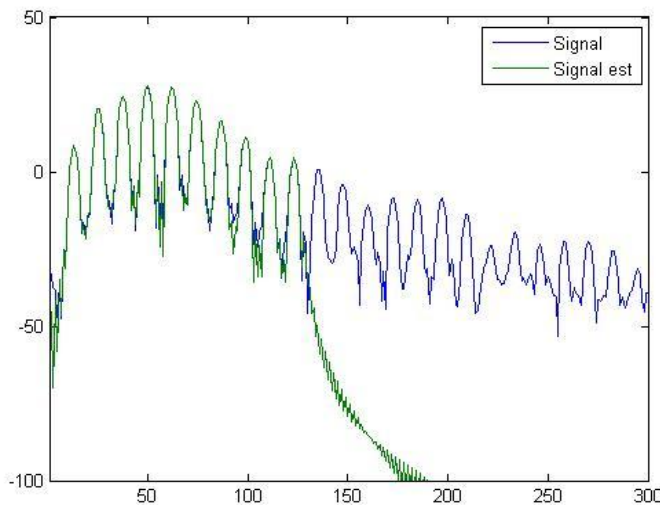
Separación de señales

- Señal de test:



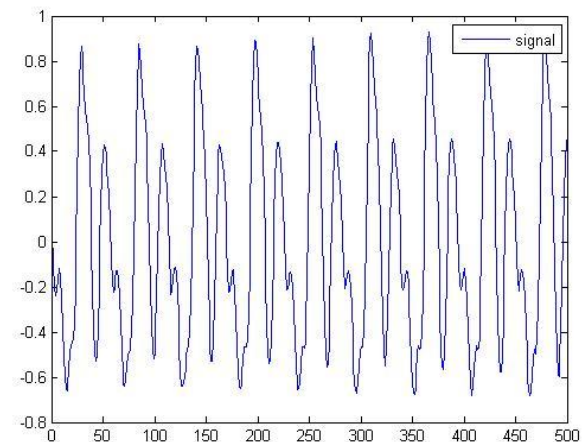
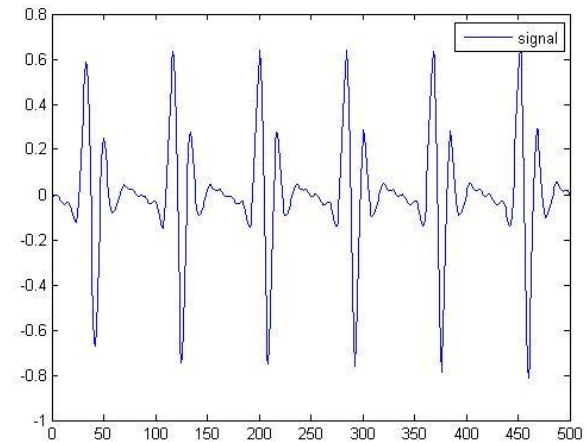
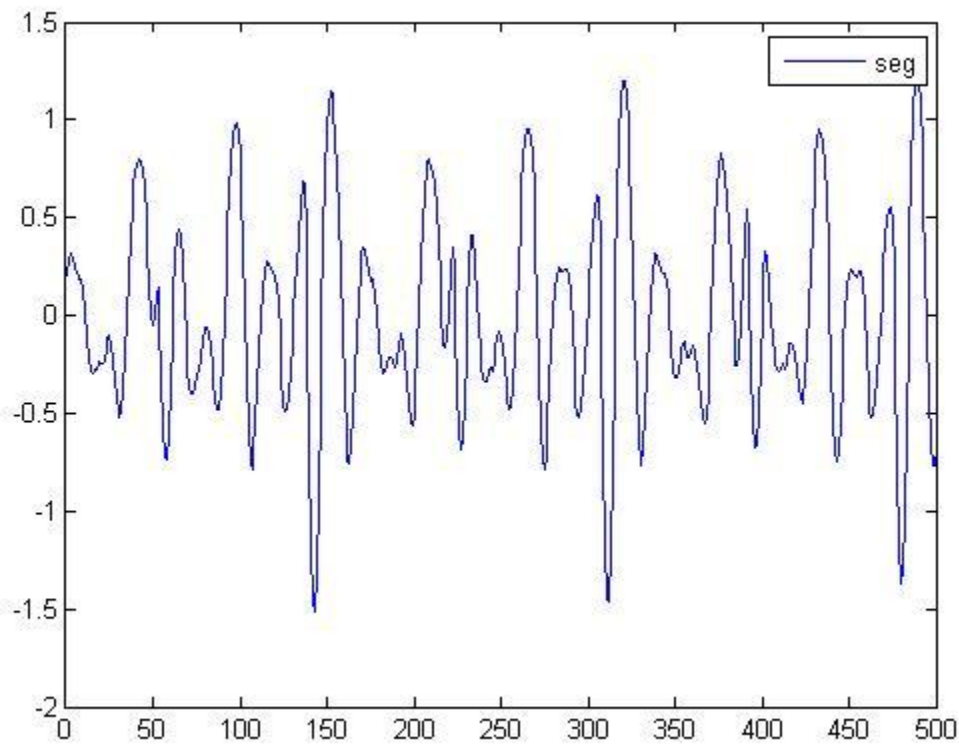
Modelado polinomial

- Procedimiento similar al explicado para señales simples.
- Es necesario estimar dos F_0 : una para cada señal.
- El resto de parámetros se obtienen por LS.
- Paso crítico: detección de F_0



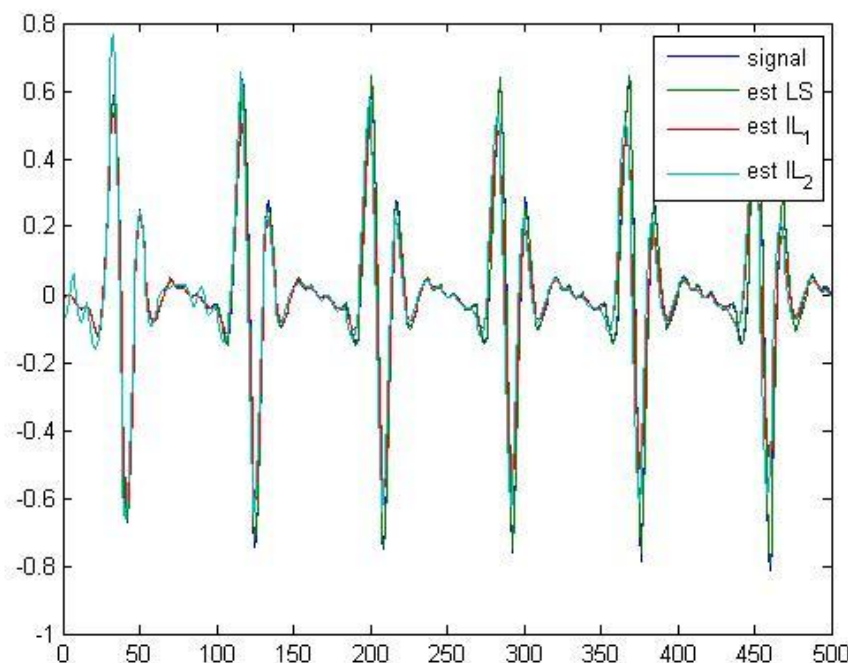
Modelado polinomial con técnicas de interpolación

$$s(n) = s_1(n) + s_2(n)$$

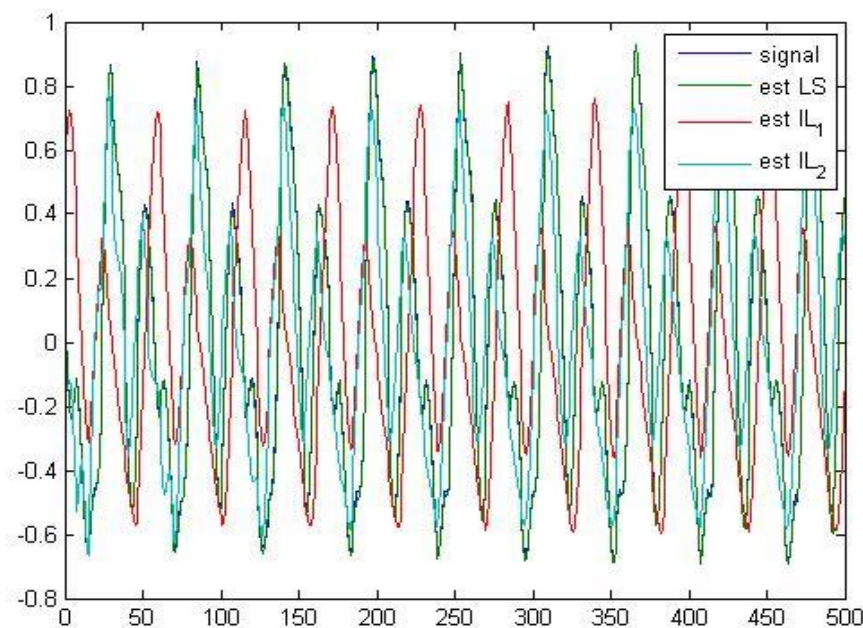


Modelado polinomial con técnicas de interpolación

S1 = Señal de Trompeta nota C4
(F0 = 523,251 Hz)



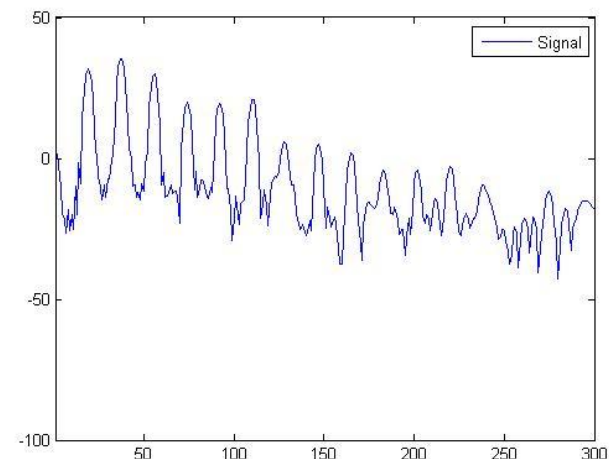
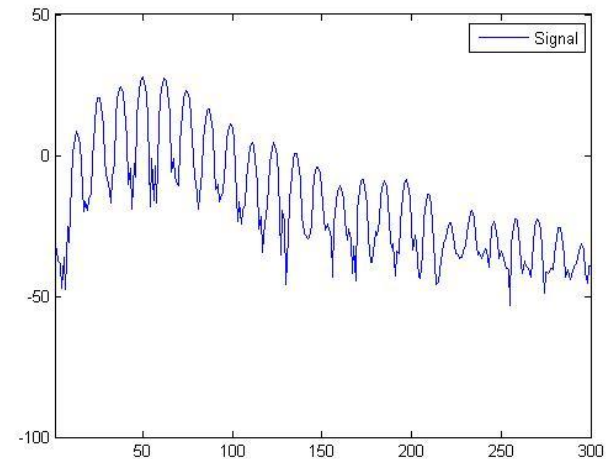
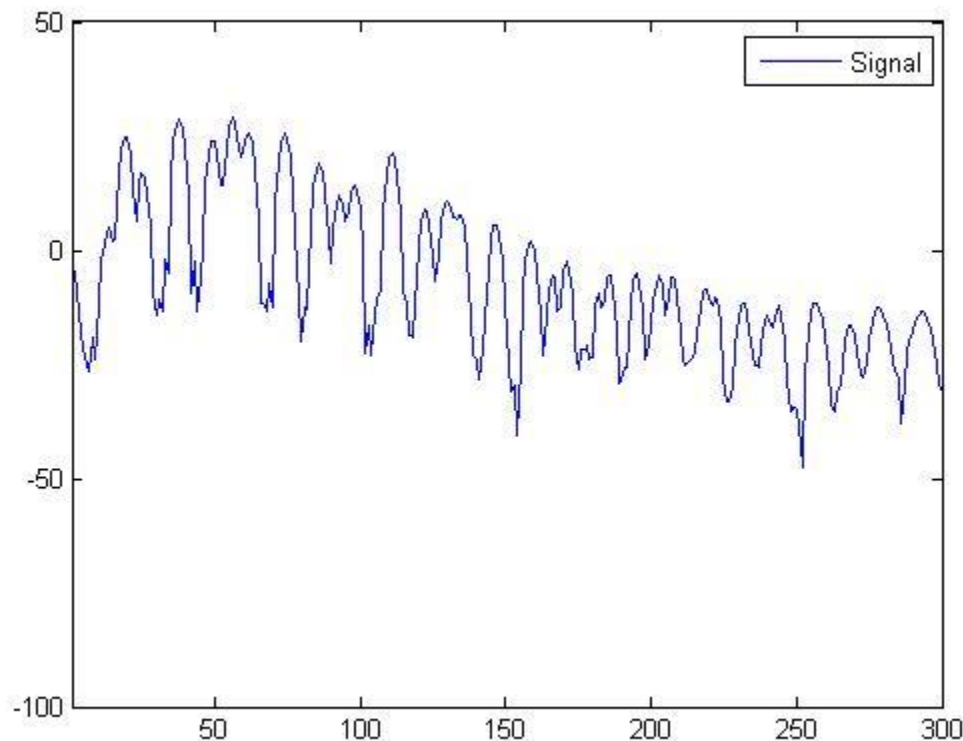
S2 = Señal de Flauta nota G4
(F0 = 783,991 Hz)



Estudio realizado para 10 armónicos, con polinomios de orden 4 y solapamiento del 50%

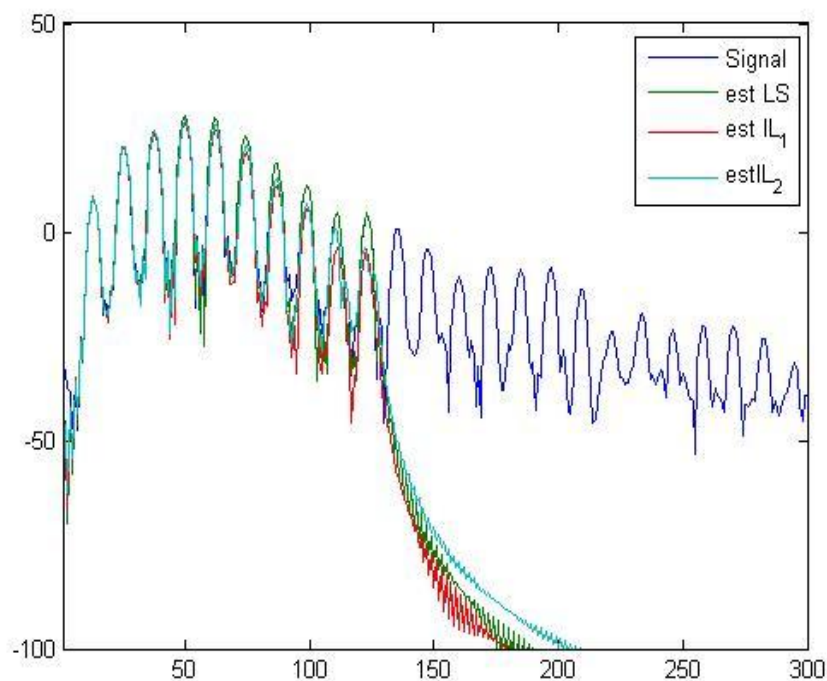
Modelado polinomial con técnicas de interpolación

$$s(n) = s_1(n) + s_2(n)$$

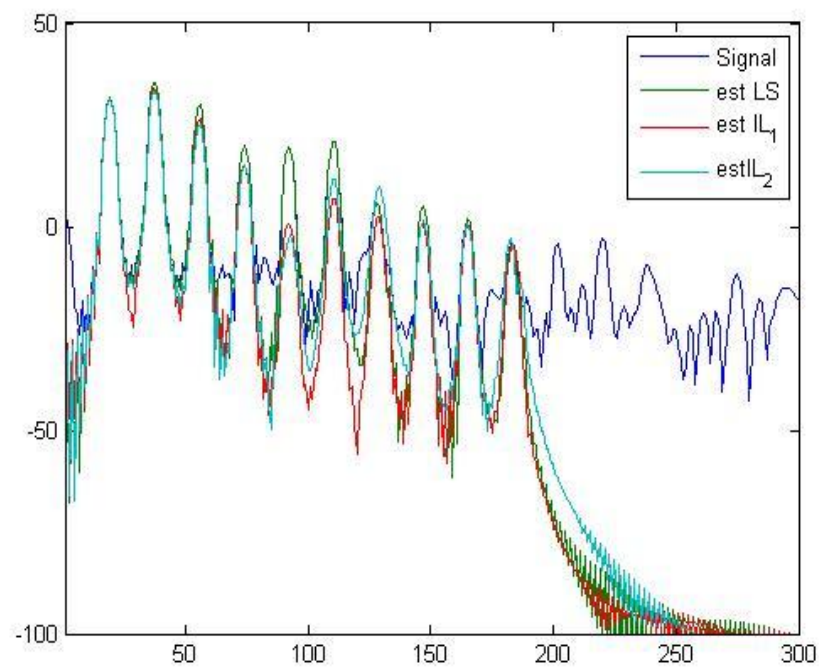


Modelado polinomial con técnicas de interpolación

S1 = Señal de Trompeta nota C4
(F0 = 523,251 Hz)



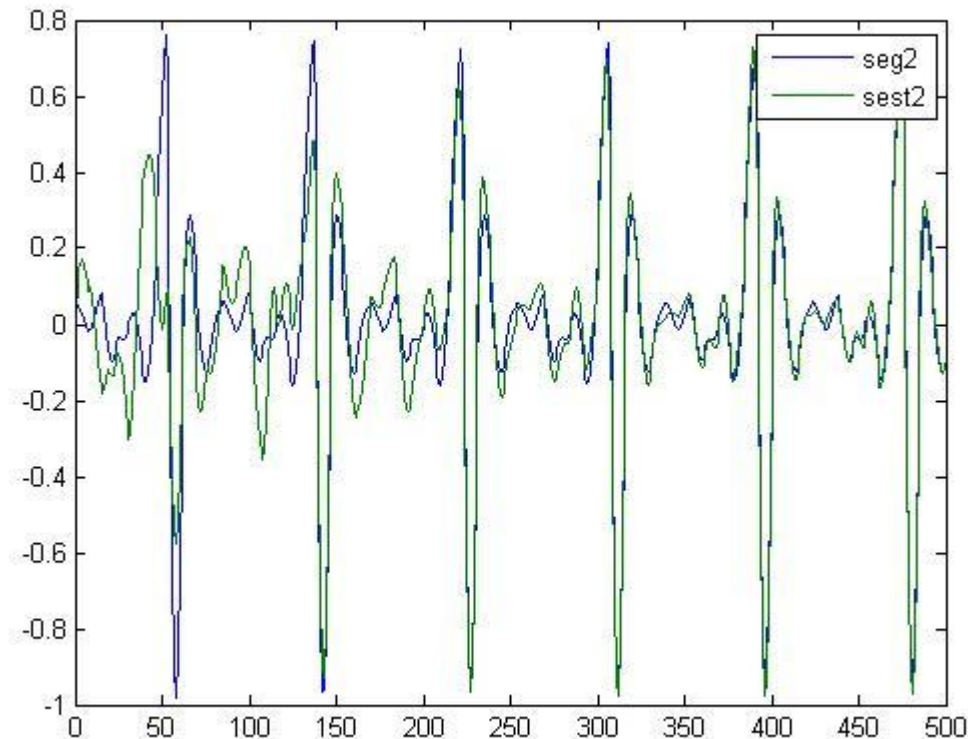
S2 = Señal de Flauta nota G4
(F0 = 783,991 Hz)



Filtro de Kalman

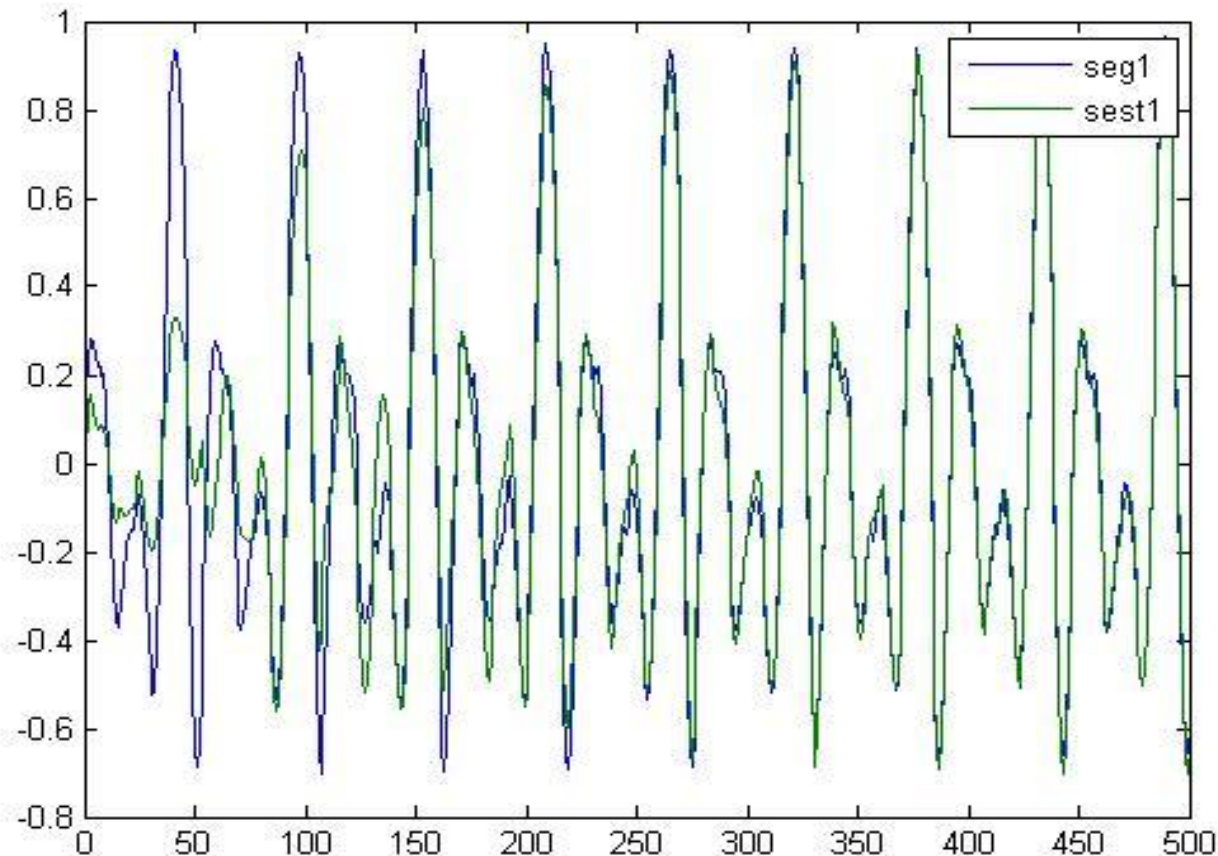
- Procedimiento similar al explicado para señales simples.
- Es necesario estimar dos F_0 : una para cada señal.

S1 = Señal de Trompeta
 nota C4
 ($F_0 = 523,251$ Hz)



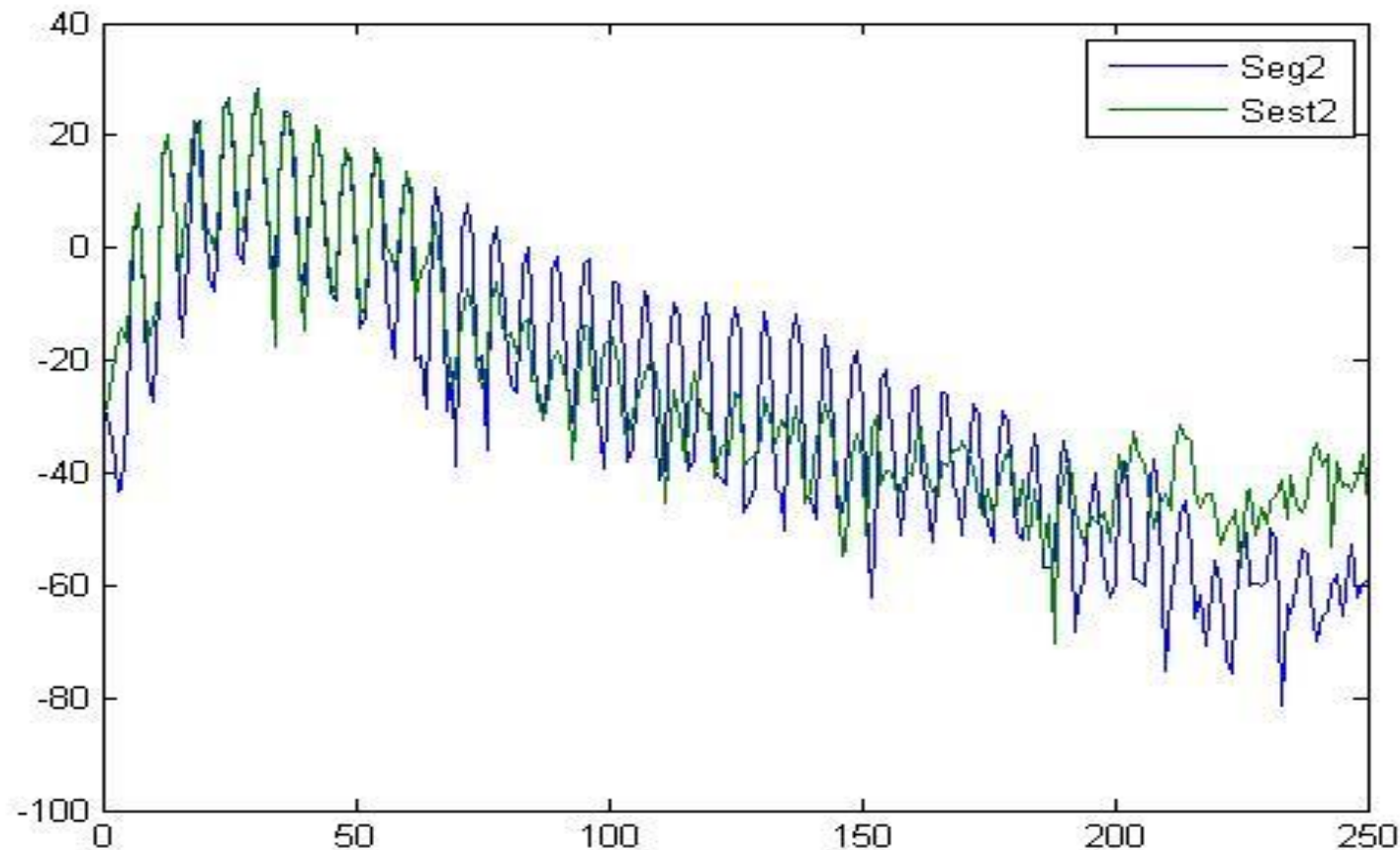
Filtro de Kalman

- S2 = Señal de Flauta nota G4 (F0 = 783,991 Hz)



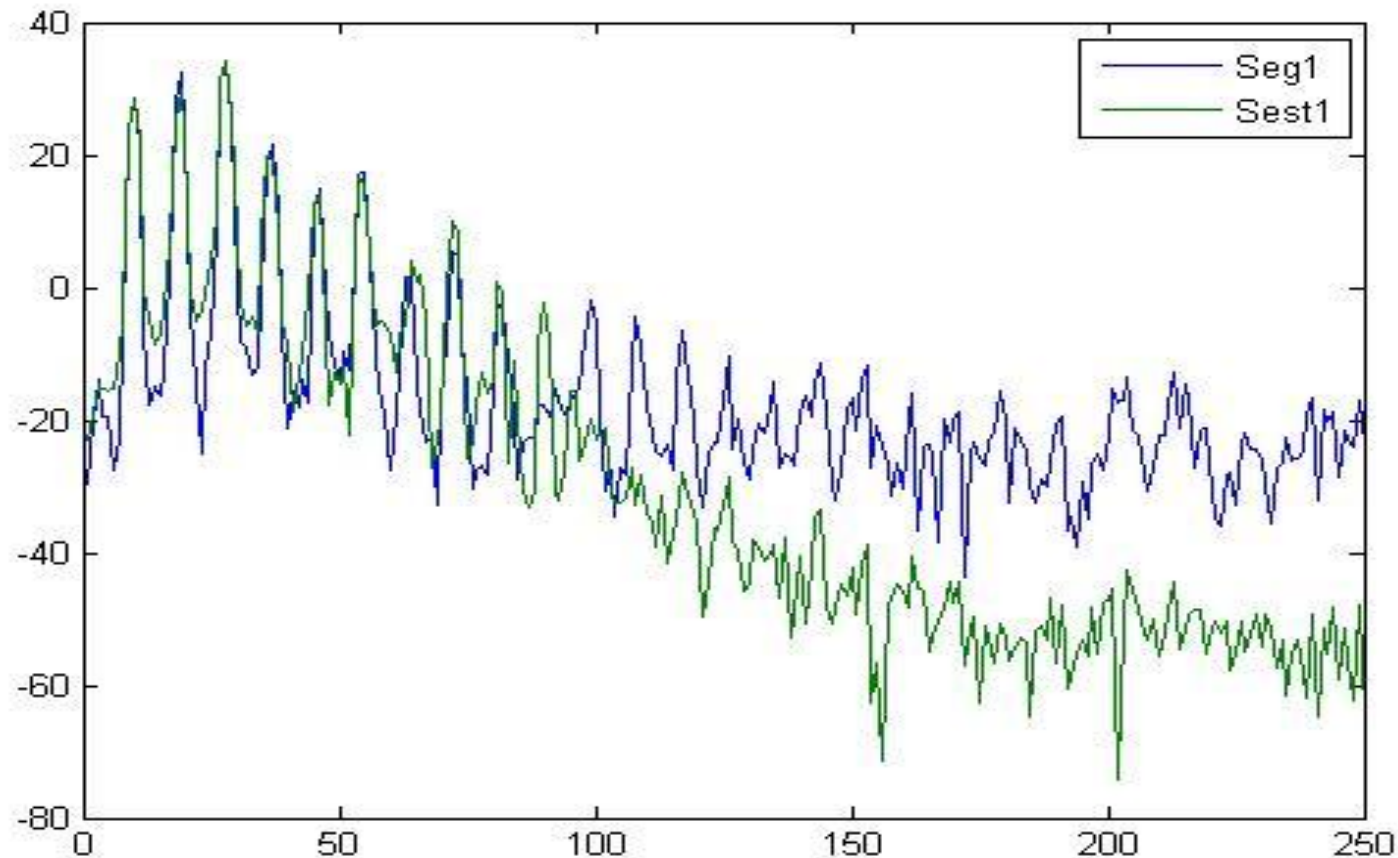
Filtro de Kalman

- Espectro de S1 = Señal de Trompeta nota C4 ($F_0 = 523,251$ Hz)



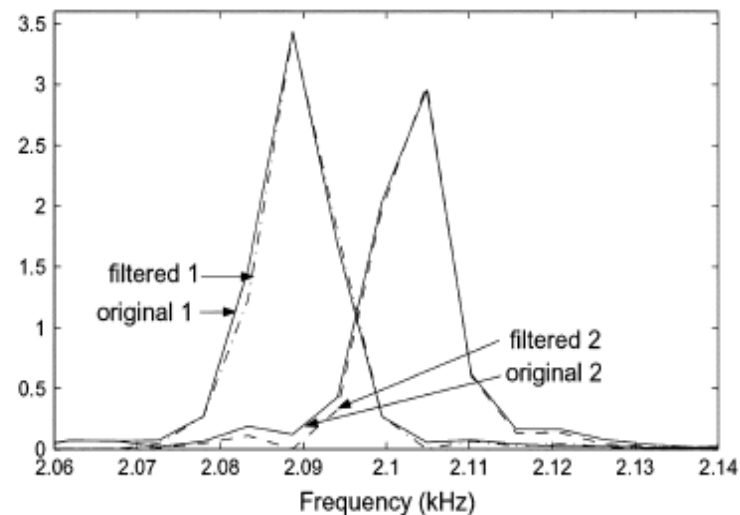
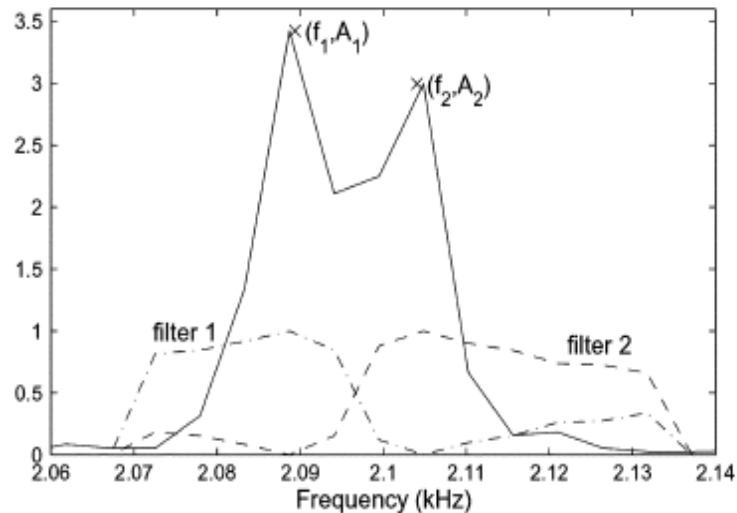
Filtro de Kalman

- Espectro de S2 = Señal de Flauta nota G4 ($F_0 = 783,991$ Hz)



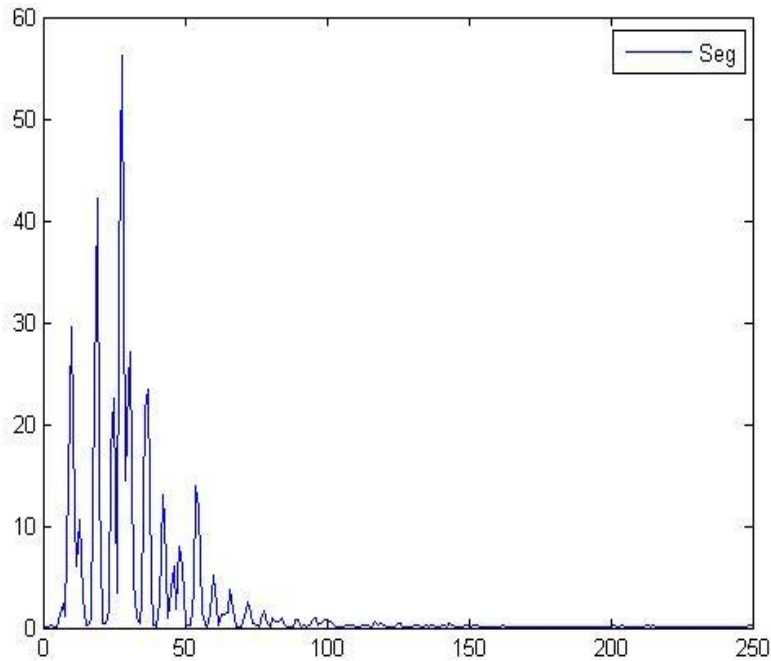
Filtrado en frecuencia

- Procedimiento desarrollado por Mark R. Every y John E. Szymanski , publicado por el IEEE en 2006.
- Utiliza el filtrado en frecuencia para dividir el pico espectral donde se produce el solapamiento y así asignar a cada fuente su parte correspondiente:

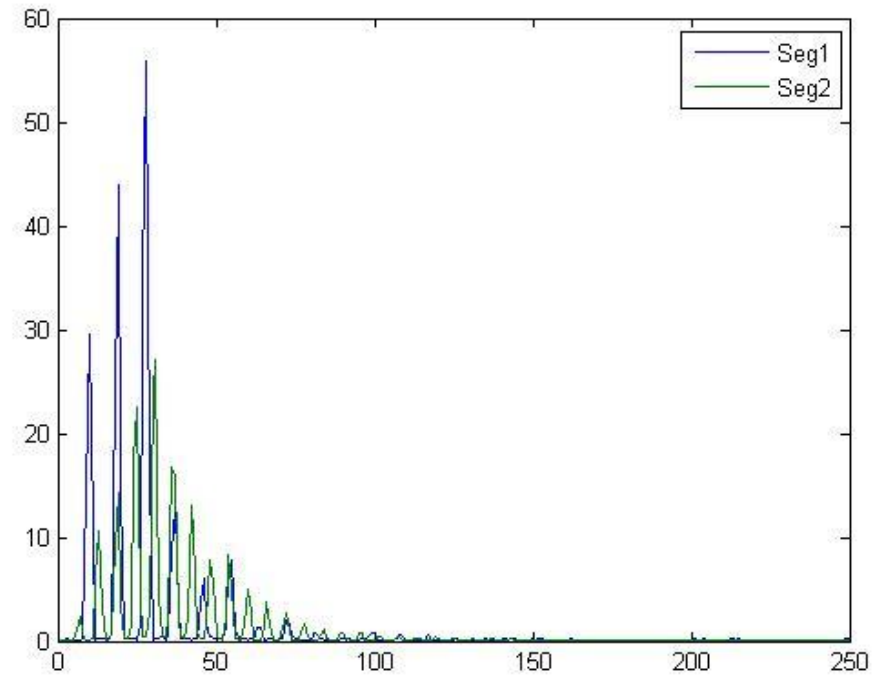


Filtrado en frecuencia

Espectro señal de test

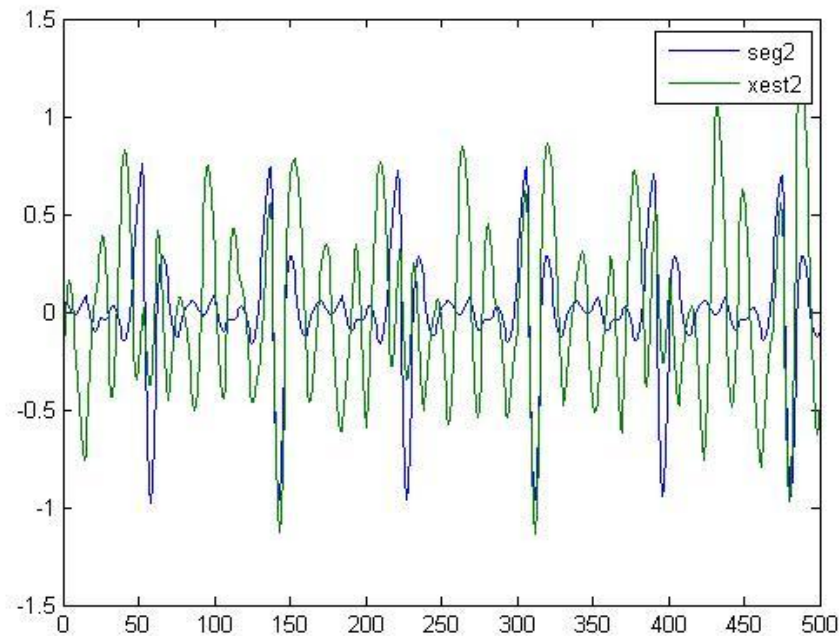


Separación de señales:

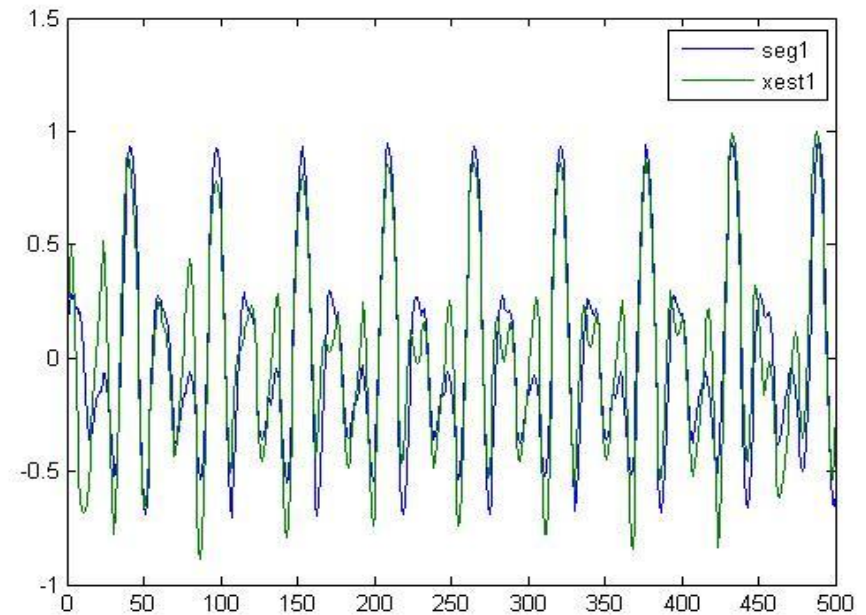


Filtrado en frecuencia

S1 = Señal de Trompeta nota C4
 (F0 = 523,251 Hz)



S2 = Señal de Flauta nota G4
 (F0 = 783,991 Hz)



Estudio comparativo

- Comparación de los métodos estudiados, utilizando como medida de calidad SRR:

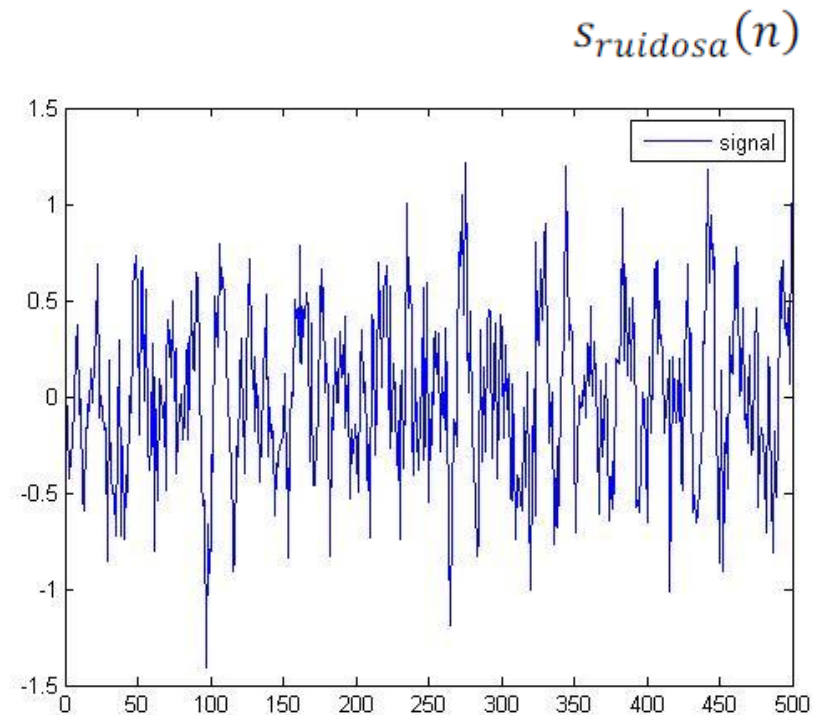
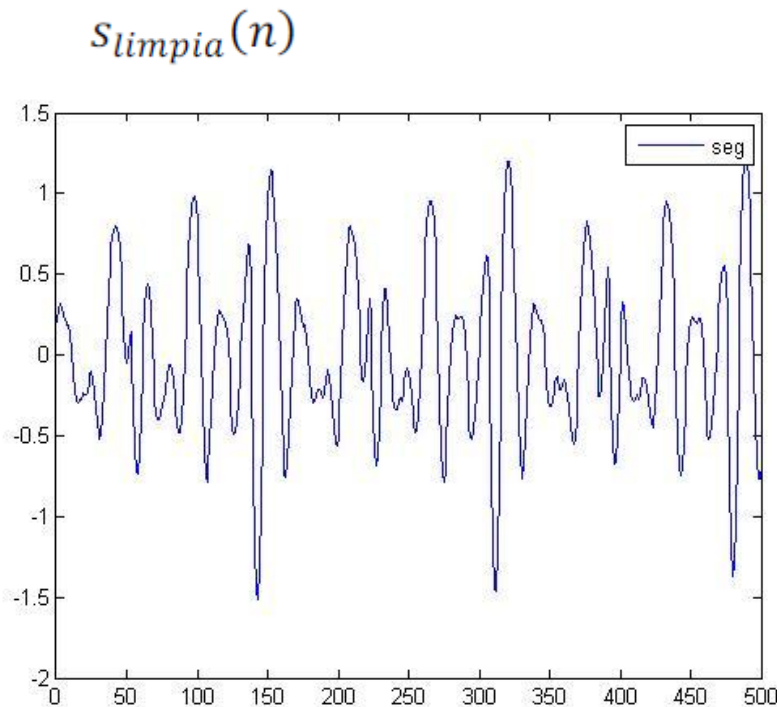
$$SRR (dB) = 10 \log_{10} \left(\frac{\sum_n s^2(n)}{E_R} \right)$$

Donde $E_R = \sum_n |s_1(n) - \widehat{s_1(n)}|^2 + \sum_n |s_2(n) - \widehat{s_2(n)}|^2$

- Estudiaremos:
 - Impacto del ruido
 - Calidad en función de la desviación de F_0

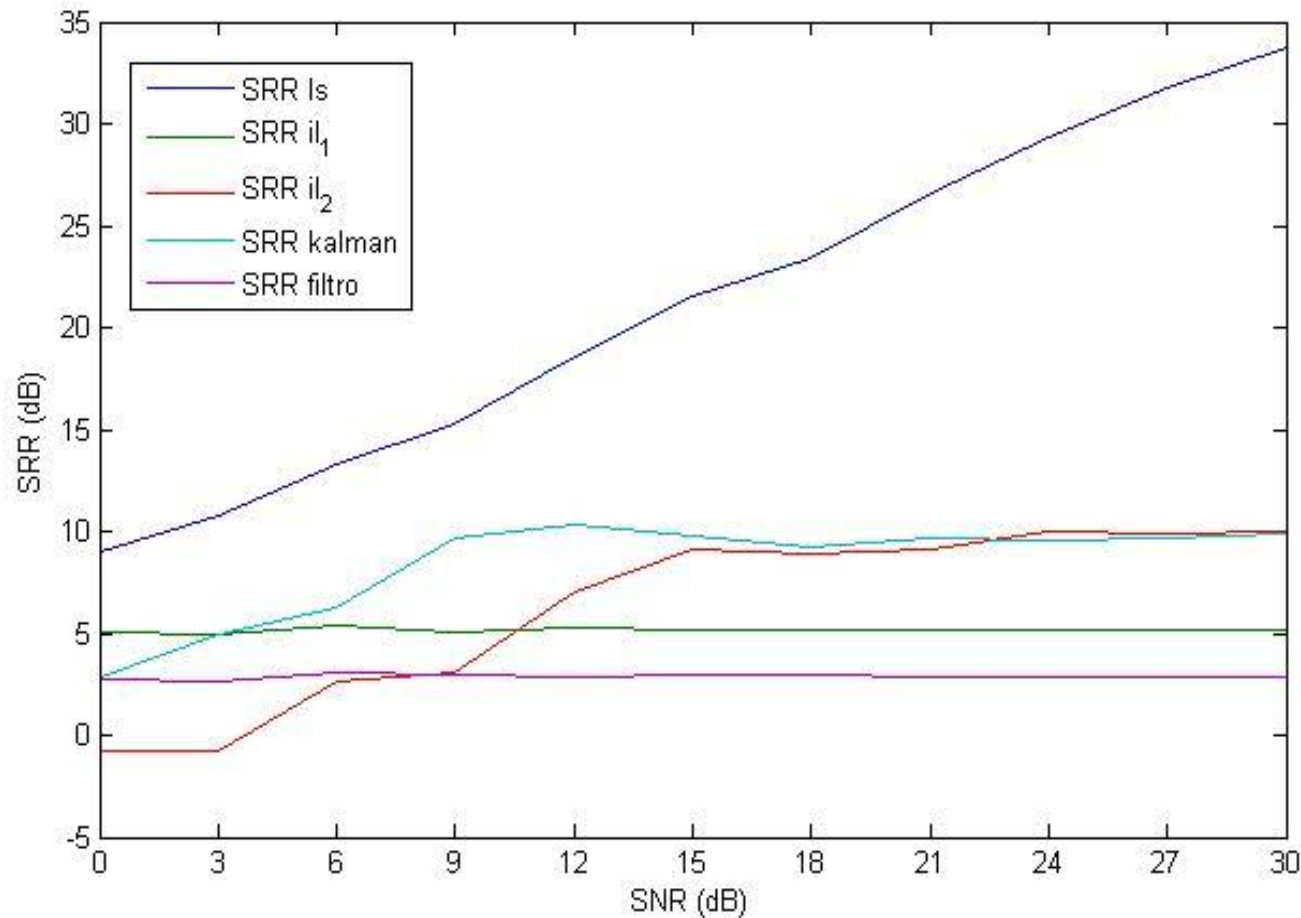
Impacto del ruido

$$s_{ruidosa}(n) = s_{limpia}(n) + r(n)$$



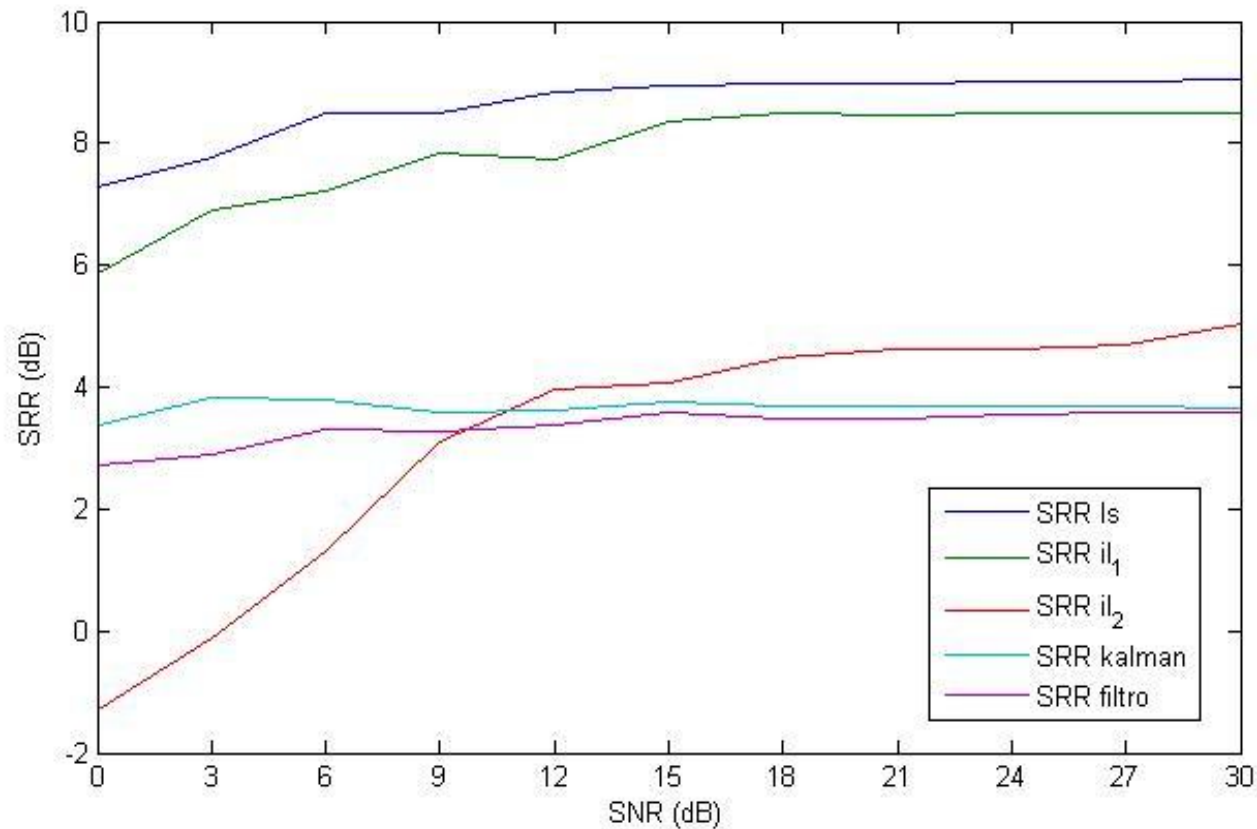
Obteniendo valores de SNR comprendidos entre 0 y 30 dB

Impacto del ruido



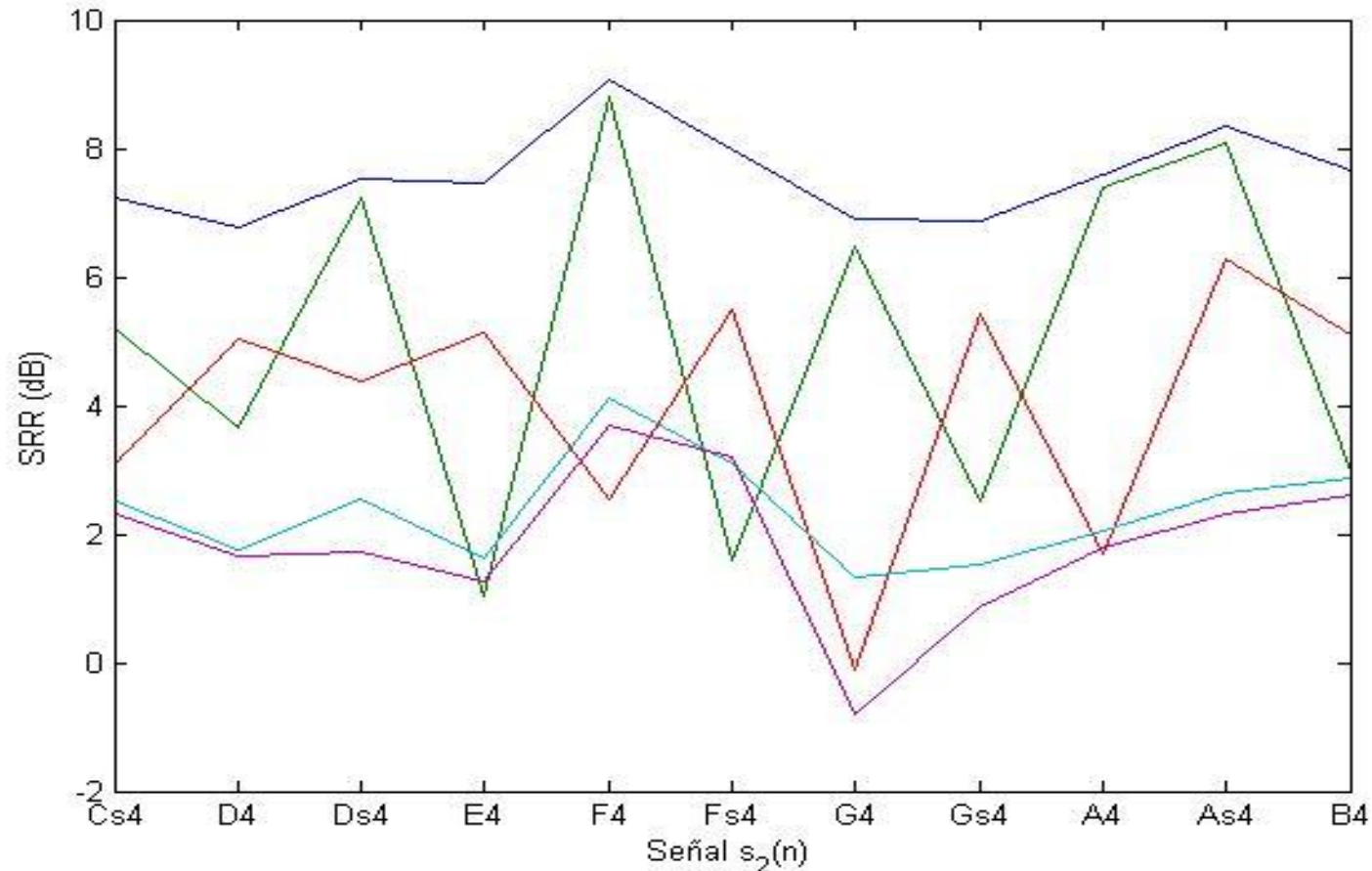
Estudio realizado para 10 armónicos, con polinomios de orden 4 y solapamiento del 50%

Impacto del ruido



(Método de detección de F_0 'Grandke')

Estudio comparativo



Estudio realizado para un solo armónico, en condiciones de SNR = 30 dB

Conclusiones

- Hemos realizado una **comparación** del funcionamiento de distintos métodos de separación de señales.
 - Mejor método: Modelado polinomial
 - Peor método: Filtrado en frecuencia
- En general, los algoritmos existentes tienen un rendimiento limitado, especialmente en el caso de **fuentes cuasi-armónicas**.

Conclusiones

- Hemos descubierto a **qué se deben las limitaciones** generales y particulares de cada método:
 - Limitaciones generales:
 - Detección de F_0
 - Limitaciones del filtro de Kalman:
 - Error en el periodo de enganche
 - Elección de valores iniciales
 - Limitaciones del filtrado en frecuencia:
 - No tiene en cuenta las propiedades generales de las señales musicales

Líneas futuras

- Hemos intentado **encontrar soluciones** para mejorar las limitaciones mencionadas, pero no se han encontrado resultados realmente satisfactorios.
- Las posibles **líneas futuras** podrían continuar por el camino que hemos dejado:
 - Modelado polinomial con técnicas de **interpolación**:
 - Utilizar otros tipos de interpolación
 - Utilizar la interpolación como condiciones restrictivas del sistema de ecuaciones, no como solución
 - Filtro de Kalman:
 - Buscar cómo evitar el error del periodo de enganche
 - Encontrar valores iniciales óptimos para los parámetros